UNIVERSITÉ DE MONTRÉAL

BEHAVIOURAL CONSIDERATIONS IN ROUTE CHOICE MODELLING

HAMZEH ALIZADEH ALIABADI

DÉPARTEMENT DES GÉNIES CIVIL, GÉOLOGIQUE ET DES MINES

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION

DU DIPLÔME DE PHILOSOPIAE DOCTOR

(GÉNIE CIVIL)

AVRIL 2018

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée :

BEHAVIOURAL CONSIDERATIONS IN ROUTE CHOICE MODELLING

présentée par : ALIZADEH ALIABADI Hamzeh

en vue de l'obtention du diplôme de : Philosophiae Doctor

a été dûment acceptée par le jury d'examen constitué de :

M. TRÉPANIER Martin, Ph. D., président

M. SAUNIER Nicolas, Ph. D., membre et directeur de recherche

Mme MORENCY Catherine, Ph. D., membre et codirectrice de recherche

M. FAROOQ Bilal, Ph. D., membre et codirecteur de recherche

M. PRATO Carlo Giacomo, Ph. D., membre

Mme FREJINGER Emma, Ph. D., membre

DEDICATION

To my dear mother Mehri, for her unconditional love and prayers…

To my dear father Ali, for his love, support and patience…

To my dear sister and brother, Hoda and Hamed for their love and encouragements…

To my dear uncle Mehdi, for his love and invaluable support…

And to my dear Nazak for her tender-loving care…

ACKNOWLEDGEMENTS

"O bearer, give me a cup, since the hidden hand that writes…

Has plans, access to which nobody can gain…

The artist of firmament, earth and spheres…

Nobody knows what plans it would feign…"

"Hafiz"

To My Advisors and Mentors…

First and foremost, I would like to express my deepest gratitude to my thesis supervisors and mentors, Professor Nicolas Saunier, Professor Catherine Morency and Professor Bilal Farooq for their invaluable guidance, support, confidence, accessibility, patience and encouragement, throughout the course of my Ph.D. at Polytechnique. You have been an incredible source of inspiration and an exceptional example of dedication. You guided me on the right path and for that I will always be thankful.

I am also very grateful to the funding sources that made my Ph.D. work possible. I was funded by the Mobility Chair of Polytechnique Montreal, headed by Professor Catherine Morency and supported by the Ministère des transports, de la mobilité durable et de l'électrification des transports du Québec (MTMDET), the Ville de Montréal, the Société de transport de Montréal (STM), and the Agence métropolitaine de transport (AMT). I also want to express my gratitude to the Mobility Chair for providing a vivid, energetic, and inspiring research environment.

I would also like to thank all the members of my Ph.D. defense committee, Prof. Carlo Prato, Prof. Emma Frejinger and Prof. Martin Trépanier for offering their valuable time and for providing thoughtful comments and suggestions to improve this thesis.

To My Friends and Colleagues…

I was lucky to pursue my research within an enthusiastic environment with friendly peers. I would like to thank all my colleagues in the transportation research group at Polytechnique Montreal. Completing this work would have been a lot more difficult were it not for your support and friendship. You have inspired and motivated me during difficult times when I needed words of encouragement. Thank you for all your support and useful advices.

To My Family…

This journey would not have been possible without the support of my mother and father who also lived all of the ups and downs of my research. Mom! Dad! You believed in me when I didn’t believe in myself! You supported me in every possible way; mentally, emotionally and financially. You have encouraged me in all my pursuits and inspired me to follow my dreams! I have always been indebted to you for all the selfless love, care, pain and sacrifice you did to shape my personality and my life, and I will always be! I’m speaking from the bottom of my hearth when I say: if I could ever be half the person you are, it would be my greatest achievement!

My dear sister and brother, Hoda and Hamed! All these years, not being by your side was the heaviest burden to bear! Not a single day goes by that I do not think of you! Nothing compares to the joy of having you in my life. Thank you for always being there beside me through thick and thin.

My dear uncle Mehdi! I would not be even close to where I am today without the encouragement, advice, and support that you've given me! I will never forget your weekly calls and our long conversations. I have always looked up to you. You are a blessing in my life. I only hope that one day I'll be able to do something equally valuable for you.

Last but not least, my heart felt regard goes to a very special person, my dear Nazak! Thank you for your love, support and understanding that not only made the completion of this thesis possible, but also enjoyable and memorable. You were always there beside me, during happy and hard moments, and motivated me by your delightful smile and your genuine belief in me.

RÉSUMÉ

La modélisation de choix d'itinéraires, i.e. de la route empruntée par les individus entre une paire origine et destination, est probablement l'un des problèmes les plus complexes et laborieux de l'analyse des comportements de déplacement. L'objectif principal de cette thèse est d'améliorer la compréhension comportementale du choix des itinéraires routiers en observant le processus sous-jacent de prise de décision des conducteurs.

Différentes approches ont été adoptées pour modéliser le comportement de choix d'itinéraires par les conducteurs, parmi lesquelles l'approche de l'utilité aléatoire et des choix discrets qui a reçu une attention considérable. Dans cette approche, chaque alternative est caractérisée par une fonction appelée « l'utilité », et les individus sont considérés comme étant des décideurs parfaitement rationnels qui maximisent leur utilité. Cette étude est basée sur une approche de modélisation en deux étapes avec échantillonnage : la première étape consiste à définir un ensemble de choix à partir duquel, en deuxième étape le choix final est fait. L'application de cette approche dans la modélisation de choix d'itinéraires routiers suscite un enjeu particulier, c'est-à-dire la définition d'ensembles de choix réalistes et représentatifs. Un autre défi de la modélisation de choix d'itinéraires est la structure complexe de corrélation entre les alternatives routières.

Bien que plusieurs approches aient été proposées pour relever les défis mentionnés ci-dessus dans la modélisation des choix d'itinéraires, l'un des défis subsistants est la cohérence des approches de modélisation proposées avec le processus comportemental sous-jacent des décisions. À cet égard, les principales contributions de cette thèse s'articulent autour de trois composantes générales de la modélisation du choix d'itinéraires en deux étapes, c'est-à-dire l'approche de modélisation, la méthode de collecte de données et l'ensemble des alternatives considérées. Les trois contributions principales concernant la première composante de la modélisation du choix d'itinéraires, c'est à dire l'approche de modélisation sont les suivantes:

- Tout d'abord, nous avons suivi l'idée que les conducteurs considèrent une représentation hiérarchique de l'espace et que certaines caractéristiques importantes de la route, c'est-à-dire les points d'ancrage, peuvent affecter leurs décisions. Nous avons étudié l'influence des ponts, comme étant des points d'ancrage, sur les choix d'itinéraires des conducteurs entre les îles de Montréal et Laval et nous avons adopté une approche de modélisation imbriquée pour représenter la hiérarchie spatiale et incorporer les effets des points d'ancrage et les

attributs de la route en même temps. De plus, cette approche offre la possibilité de capturer les similitudes non observées des itinéraires qui traversent un même point d'ancrage, telles que la sécurité routière, le paysage, le confort de conduite, etc. Les résultats montrent que l'approche de modélisation proposée correspond mieux aux données et surpasse les capacités de prédiction des modèles de choix comparatifs basés exclusivement sur les attributs de la route, tels que le « Path-Size Logit », le « Extended Path-Size Logit », et le « Independent Availability Logit ».

- La deuxième contribution de cette thèse repose sur l'idée que les individus ont des propensions différentes à choisir parmi les différents itinéraires disponibles entre une origine et une destination. Ceci résulte principalement de leurs différences d’attitudes, de préférences et d’expériences, et par conséquent, peut former différents types de comportements de prise de décision. À cette fin, nous avons étudié un ensemble longitudinal de données GPS, en suivant 1 746 chauffeurs de taxi effectuant plus de 22 000 déplacements sur une période d'un an. En conséquence, quatre catégories de stratégies ont été trouvées en fonction des variations observées dans les déplacements effectués pendant les jours et les nuits, et entre les déplacements de courte et longue durées. La compréhension de ces stratégies opérationnelles permet non seulement de mieux comprendre la dynamique de la circulation urbaine, ce qui est très important pour la ville et les planificateurs de transports, mais la classification comportementale offre aussi la possibilité d'estimer des modèles de choix d’itinéraires plus précis.

- La troisième contribution tourne autour de l'idée que les choix sont fortement influencés par: 1) les traits latents et les variables comportementales qui ne peuvent être directement observés et mesurés, comme les attitudes, les perceptions et les préférences de style de vie, et 2) l'hétérogénéité latente existante entre les différents segments de la population. Pour incorporer correctement l'effet de l'hétérogénéité des segments et faire la distinction entre les comportements de choix des différentes classes de la population étudiée, nous avons utilisé un modèle de classes latentes, dans lequel nous avons intégré le rôle des variables latentes. Nous avons appliqué le modèle proposé pour comparer le comportement de choix d’itinéraires des conducteurs fréquents par rapport aux conducteurs occasionnels. Les résultats confirment que les différents segments de la population étudiée se comportent différemment dans leurs choix d’itinéraires et les résultats de la modélisation démontrent que l'inclusion des traits de comportement dans le modèle de classes latentes améliore considérablement son ajustement sur les données.

Malgré l'attrait des cadres de modélisation comportementale, leurs applications dans les études de choix d'itinéraires restent rares, ce qui peut être principalement lié au fait que la collecte de données comportementales est laborieuse, couteuse et prend énormément de temps. Une autre contribution de cette thèse concerne la conception et la mise en œuvre d'un cadre de collecte de données conçu pour les études comportementales de choix d'itinéraire. L'objectif principal du cadre proposé est de recueillir des données reflétant la stochasticité des préférences des individus et la nature complexe des processus de prise de décision des conducteurs, sans augmenter de manière significative le fardeau des répondants. L'enquête recueille des informations sur les caractéristiques sociodémographiques et socio-économiques des conducteurs, leurs choix d'itinéraire et leurs ensembles d'alternatives considérées, ainsi que leurs perceptions et leurs traits de comportement. À la fin de la période de collecte de données (trois mois), 843 personnes ont participé à l'enquête, dont 539 (64 %) l'ont terminé, alors que les autres 304 (36 %) l'ont abandonné. La durée moyenne de l'enquête était d'environ 16 minutes. Nous avons également examiné diverses paradonnées disponibles, telles que le temps de complétion et le taux de décrochage par type de question et de section, quelques statistiques sur diverses méthodes de recrutement et le taux de complétion de l'enquête par heure de la journée. Ces paradonnées peuvent être utiles pour identifier les problèmes de collecte de données, proposer de nouvelles stratégies de collecte de données et déterminer un compromis entre la qualité des données, le coût et la période de collecte de données.

Le dernier domaine de contribution de cette thèse concerne la première étape de la modélisation du choix d'itinéraires dans un cadre de modélisation en deux étapes, c'est-à-dire la compréhension des ensembles de choix d'itinéraires. Étant donné que ces ensembles sont généralement latents pour l'analyste et que les informations concernant les attributs objectifs et subjectifs affectant leurs tailles et compositions sont limitées, nous avons adopté le sondage mentionné ci-dessus pour recueillir des informations sur les ensembles de choix considérés par les répondants. Nous avons étudié 988 itinéraires, déclarés par 506 conducteurs résidant et conduisant dans la grande région de Montréal. L'effet de six catégories de facteurs sur la taille des ensembles déclarés a été étudié, y compris les attributs personnels, les facteurs déclarés, les indicateurs comportementaux, les incitatifs, les déterminants de la conscience et les composantes spatiales, temporelles et environnementales. En conséquence, quatre groupes différents ont été définis en fonction du nombre d'alternatives considérées et la relation de ces facteurs avec chaque groupe a été étudiée. Les résultats de cette étude permettent de mieux saisir la relation entre divers attributs et la taille de l'ensemble de choix. Une meilleure compréhension de la taille et de la composition des ensembles de choix peut considérablement améliorer les résultats d'estimation et de prévision des modèles de choix d'itinéraires en fournissant de l'information sur les préférences des individus.

Les modèles de choix élaborés incluant les traits et variables de comportement sont plus difficiles à estimer et nécessitent des méthodes de collecte de données plus complexes. Cependant, ces modèles offrent de meilleures capacités de prédiction, corrigent les biais cognitifs (i.e. erreurs de perception, d'évaluation, et d'interprétation logique), vérifient les hypothèses comportementales concernant les processus de prise de décision et, par conséquent, fournissent une référence pour évaluer la performance des modèles plus parcimonieux.

ABSTRACT

Route choice modelling is probably one of the most complex and challenging problems in travel behaviour analysis. It investigates the process of route selection by an individual, making a trip between predefined origin-destination pairs. The main objective of this thesis is to enhance the behavioural understanding of drivers' route choice decisions by observing drivers' underlying process of decision-making.

Various approaches have been adopted to model drivers' route choice behaviour, among which random utility discrete choice models have received considerable attention. In this modelling framework, each choice alternative is characterized by a function called utility, and individuals are viewed as rational decision makers who maximize their perceived utilities. This work is based on a two-stage random utility maximization framework with sampling of alternatives, in which the first stage consists of defining a proper consideration set, from which the final choice is made in the second stage. The application of the two-stage random utility maximization framework in route choice modelling gives rise to a particular challenge, namely the definition of realistic and representative choice sets. Another challenge of route choices modelling is the complex correlation structure of route alternative. Although several approaches have been proposed to tackle the above-mentioned issues, one of the remaining challenges is the consistency of the proposed modelling approaches with drivers' underlying behavioural process of decision-making.

In this regard, the main contributions of this thesis revolve around the three general components of the two-stage route choice modelling framework, namely the modelling framework, the data collection method, and the consideration set of route alternatives. The three main contributions regarding the modelling framework include:

- First, we followed the idea that drivers follow the hierarchical representation of space and that some prominent features of the route, i.e. anchor points, might affect their decisions. We studied the influence of bridges as anchor points on drivers' route choice decisions between the two islands of Montreal and Laval, and adopted a nested modelling approach to represent the space hierarchy and to incorporate the effects of anchor points and route level attributes at the same time. Moreover, this approach also provides the possibility to capture the unobserved similarities of routes crossing a same anchor point, such as safety, scenery, driving comfort, etc. Results show that the nested modelling approach provides

better model fits and outperform the prediction abilities of comparative route-based choice models, such as the Path-Size Logit, Extended Path-Size Logit, and Independent Availability Logit models.

- The second contribution of this thesis builds upon the idea that individuals have different inclinations towards choosing a route between an origin and destination, which stems from having different attitudes, preferences, and experiences, and consequently may form different types of decision-making behaviours. For this purpose, we studied a longitudinal GPS dataset, tracking 1,746 taxi drivers making more than 22,000 trips over a period of one year. Accordingly, four categories of operating strategies have been found based on variations in trips made during days and nights, and between short trips and long trips. Apart from the fact that understanding of these operating strategies helps to better understand urban traffic dynamics, which is very important to the city and transportation planners, the behavioural classification provides the possibility of estimating more accurate route choice models.
- The third contribution revolves around the idea that choices are greatly influenced by: 1) latent traits and variables that cannot be directly observed, such as attitudes, perceptions, and lifestyle preferences, and 2) the latent heterogeneity existing between different segments of the population. To properly incorporate the effect of segment heterogeneity and to distinguish between choice behaviours of different classes of the sample population we used a latent class model, in which we incorporated the role of the underlying attitudinal and behavioural traits using an Integrated Choice and Latent Variable model. We applied the proposed modelling framework to compare the route choice behaviour of frequent versus occasional drivers. Expectedly, major behavioural traits have been observed among different segments of the studied population, and modelling results demonstrated that the inclusion of behavioural traits in the LC model significantly improves its fit over the data.

Despite the appeal of behavioural modelling frameworks, their applications in route choice studies remain rare, which can be mostly related to the fact that collecting behavioural data is cumbersome, costly and time consuming. Another contribution of this thesis is the design and implementation of a data collection framework designed for behavioural route choice studies. The main aim of the proposed framework is to collect data reflecting the stochasticity of individuals' preferences and the complex nature of the drivers' decision-making processes, without significantly increasing the respondent burden. The survey collects information on drivers' sociodemographic and socioeconomic characteristics, their revealed route choices and their considered sets of route alternatives, as well as their perceptions, and behavioural traits. By the end of the three-month data collection period, 843 individuals started the survey from which 539 (64 %) completed it, while the remaining 304 (36 %) dropped out at various points of the survey. The average completion time of the survey was around 16 minutes. We also looked at various available paradata, such as the completion times and dropout rates per question type and section, some statistics on various recruitment methods, and survey completion rate per hour of the day. Looking at these paradata can be useful in identifying data collection problems, proposing new data collection strategies, and determining a trade-off between data quality, cost and time.

The last contributing area of this thesis concerns the first stage of route choice modelling in a two-stage modelling framework, namely the understanding of route choice sets. Since actual consideration sets of route alternatives are usually latent to the analyst and information concerning objective and subjective attributes affecting their size and composition is limited, we adopted the proposed survey framework to collect information on respondents considered choice sets. We investigated 988 route alternatives, declared by 506 drivers, residing and driving in the Greater Montreal Area. The effect of six broad categories of factors on the size of drivers' consideration sets has been studied, including personal attributes, declared factors, behavioural indicators, incentives, awareness determinants, and spatial, temporal and environmental components. Accordingly, four different clusters were defined based on the number of considered alternatives and the relationship of these factors with each cluster was investigated. Results of this study shed light on the relationship of various attributes with the size of the choice set. A better understanding of drivers' consideration sets' size and composition can significantly improve route choice models' estimation and prediction efficiency by providing information about travellers' preferences.

Behaviourally elaborated models require customized programs and fast computers for estimation and necessitate well-tailored data collection methods. However, these models provide better prediction abilities, correct for cognitive biases (i.e. errors of perception, evaluation, and logical interpretation), verify behavioural hypotheses regarding the decision-making process, allow for a clearer behavioural interpretation than standard choice models, and hence, provide a benchmark to evaluate the performance of more parsimonious models.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

Behvr	Behavioural traits
BSS	Between-group Sum of Squares
PB	Paper Based
CB	Computer Based
CBD	Central Business District
CCS	Considered Choice Set
CIRRELT	Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation
CNL	Cross Nested Logit
EPSL	Extended-Path-Size Logit
Fact	Affecting Factors
GIS	Geographical Information System
GMA	Greater Montreal Area
GPS	Global Positioning System
GRIP	Graphical Route Information Panel
GTHA	Greater Toronto and Hamilton Area
HAC	Hierarchical Agglomerative Clustering
HTS	Household Travel Survey
i.i.d.	Independently and Identically Distributed
IAL	Independent Availability Logit
ICLV	Integrated Choice and Latent Variable
IIA	Independence of Irrelevant Alternatives
KMO	Kaiser-Meyer-Olkin
LC	Latent Class
LK	Logit Kernel
MEV	Multivariate Extreme Value
MH	Metropolis-Hastings
MNL	Multinomial Logit Model
MRI	Mental Representation Items
MTMDET	Ministry of transport, sustainable mobility and transport electrification

MTQ	Ministry of Transportation of Quebec
NL	Nested Logit
NSERC	Natural Sciences and Engineering Council of Canada
Obs	Observed choices
OD	Origin-Destination
PCA	Principal Component Analysis
Percp	Perceptions
PFA	Principal Factor Analysis
PS	Path-Size
PSL	Path-Size Logit
RNMEV	Recursive Network Multivariate Extreme Value
RP	Revealed Preferences
RRM	Random Regret Minimization
RUM	Random Utility Maximization
SP	Stated Preferences
TAZ	Traffic Analysis Zones
TV	Test Value
WB	Web Based
WTP	Willingness to Pay

CHAPTER 1 INTRODUCTION

1.1 Context

In a transportation network, individuals' route choices or more generally, travels, are interpreted as demand. Travel demand may cause congestion when it exceeds the offered capacity, i.e. the supply. More than 84 % of Canadian households owned or leased at least one vehicle in 2007 (Natural Resources Canada, 2009). Consequently, many trips are made by vehicles, which in turn may induce road congestion. Congestion reduces our quality of life, affects our economy and degrades our environment by among others, wasting energy and producing air contaminants (Chen, 2013a; Frejinger, 2008; Hoogendoorn-Lanser, 2005; Sikka, 2012).

Congestion is a growing problem in Canada as well as in many other countries. Car dependency among Canadians is rising, while active transportation (walking or biking) is declining (Turcotte, 2008). In 2006, Transport Canada studied the annual cost of congestion in nine of the largest cities in Canada. The total annual cost has been evaluated to be $3.1 billion, with Montreal being the second contributor to this cost after the city of Toronto. Measures such as duration of the peak period, percentage of work versus non-work trips, values of time, fuel price, and the unit cost to mitigate greenhouse gases were considered in the calculation (The High Cost of Congestion in Canadian Cities, 2012). Another study conducted by Metrolinx in 2008, investigated the cost of congestion based on the difference between an optimal speed and an actual speed during morning and evening rush hours, and concluded that the annual social and economic cost of congestion in the Greater Toronto and Hamilton Area (GTHA) was approximately $3.3 billion (Metrolinx, 2008). Another study by the Ministry of Transportation of Quebec – MTQ (today known as the ministry of transport, sustainable mobility and transport electrification – MTMDET) evaluated the cost of congestion in Montreal to be around $1.4 billion per year, considering a traffic speed threshold of 60 %. Measures such as time, vehicle wear, fuel, pollution and greenhouse gas emissions contributed to the estimated cost (Ministère des Transports du Québec, 2009).

Travel is essential to the social welfare and a healthy economy. To alleviate the adverse effects of traffic congestion, it is imperative, first of all, to understand its causes. One of the major causes of congestion is the excess of demand over supply. Road demand, which materializes as traffic flow patterns, is in turn the result of individuals' route choice decisions. Therefore, it is of critical importance to understand drivers' travel and route choice behaviour in order to mitigate the adverse effects of congestion. Route choice modelling is very important in transportation planning and is a powerful tool to forecast traffic flow patterns, design new transportation infrastructures, and investigate new transportation policies.

1.2 Background

Route choice models investigate the process of route selection by an individual, making a trip between given origin-destination (OD) pairs. There is a large body of literature in microeconomics, behavioural science, psychology, and behavioural geography focusing on the improvement of the understanding of the underlying process of decision-making. Accordingly, several modelling frameworks have been proposed to simulate drivers' route choice behaviour. For instance, prospect theory (Gao, Frejinger, & Ben-Akiva, 2010; Kahneman & Tversky, 1979) and cumulative prospect theory (Connors & Sumalee, 2009; Tversky & Kahneman, 1992; Xu, Zhou, & Xu, 2011) have been applied by researchers to take into account the limited rationality of drivers in making decisions, by incorporating psychological and behavioural aspects. The uncertainty and imprecision of drivers in making route choice decisions have been considered in a Fuzzy Logic modelling framework (Henn, 2003; Luisa De Maio & Vitetta, 2015; Murat & Uludag, 2008; Quattrone & Vitetta, 2011), and artificial neural networks have been used to account for the non-linearity of the decision-making process by imitating the human conscious structure (Dougherty, 1995; Kim, Sung, Namgung, & Jang, 2005). Also, the Random Regret Minimization (RRM) approach has been adopted by Prato (2014) in a route choice modelling context, in which choice makers tend to choose the alternative that minimizes the regret of not having chosen the other alternatives.

Among the proposed approaches, random utility discrete choice models are among the most frequently used to model, analyze and understand decision-making behaviours (Prato, 2009b; Walker, 2001). This approach presumes that decision makers tend to choose the best alternative by maximizing a specific perceived utility. In other words, in a random utility maximization approach, observed choices are manifestations of decision makers' preferences, expressed by alternative specific utility functions. Choice model estimation results in selection probabilities for each alternative, which can then be used to predict decision makers' choice behaviour (Dhaker, 2012; Frejinger, Bierlaire, & Ben-Akiva, 2009; Manski, 1977; Walker, 2001). Since the decision maker may not have a perfect knowledge and because the effect of some attributes cannot be directly measured (such as lifestyle and attitudes), an error component has been introduced into the utility function to take into account the stochasticity and imprecision caused by decision makers' uncertainty and behavioural randomness (Ben-Akiva & Bierlaire, 2003).

In this framework, a two-stage modelling process with sampling of alternatives is mostly adopted to simulate drivers' route choices. In the first stage, decision makers form a limited set of route alternatives, called the consideration set, from the universal set, i.e. all the possible routes between the studied OD pair. Then, they pick their most preferred routes (based on a specified utility function) from the consideration set in the second stage (Ben-Akiva & Boccara, 1995; Bovy, 2009; Manski, 1977). This thesis is built upon the two-stage random utility maximization framework with sampling of alternatives (described in more details in subsection 2.1.1).

1.3 Motivation

Route choice modelling is probably one of the most complex and challenging problems in traffic assignment. In a two-stage framework, the complexity of modelling route choice decisions is mainly attributed to the high density of the road network and the large number of possible alternatives between OD pairs.

In a two stage modelling approach with sampling of alternatives, defining a proper consideration set is a serious challenge in route choice modelling. Consideration sets of route alternatives are rarely observed and are usually latent to the analyst. Therefore, deterministic and stochastic path generation techniques are usually adopted to generate simulated sets of considered alternatives, using variations of the shortest path algorithm (Ben-Akiva & Boccara, 1995; Bovy, 2009; Prato, Bekhor, & Pronello, 2012). The generated sets should include alternatives that are attractive to the driver in a real world choice situation (Hess & Daly, 2010), and the misspecification of their size and composition greatly affects model's estimates and may lead to fallacious predicted demand levels (Bliemer & Bovy, 2008; Geda, 2014; Peters, Adamowicz, & Boxall, 1995; Prato & Bekhor, 2006, 2007b; Schuessler & Axhausen, 2009; Swait & Ben-Akiva, 1987a).

Another major challenge in route choice modelling is to take into consideration the correlation structure (i.e. overlaps) among various routes. Several approaches have been proposed to address this issue, either by modifying the deterministic part of the utility function or its stochastic part (Dhaker, 2012; Prato, 2009a). These approaches are discussed in detail in Chapter 2.

Moreover, the general complexity of choice modelling, which is also extended to route choice modelling, is related to factors such as the sophisticated nature of human behaviour, the ambiguity of the decision-making process, and the stochasticity of individuals' preferences. The heterogeneity in travellers' behavioural characteristics, in conjunction with the complex effect of route attributes, further increases the inherent complexity of route choice modelling.

Individuals have different inclinations towards choosing a route between an origin and destination. This heterogeneity comes from having different preferences, different experiences, and different attitudes. For instance, in terms of preferences, an individual might prefer to choose a route with better scenery while another individual seeks the fastest route; in terms of experiences, more experienced drivers may be more familiar with the network; and in terms of attitudes, some drivers might have better spatial abilities.

Drivers' attitudes, perceptions and preferences play a major role in the decision-making process (Ben-Akiva et al., 2002; Prato et al., 2012; Walker, 2001). However, the effect of these factors on drivers' route choice decisions is not easy to evaluate using conventional methods of data collection. Route choice studies are mostly based on GPS data and revealed preference surveys, where there is a notable lack of behavioural and attitudinal information. Moreover, stated preference surveys usually avoid attitudinal questions to minimize respondents' burden and maximize surveys' completion rate.

1.4 Objectives and Contributions

Although several approaches have been proposed to tackle the above-mentioned issues in route choice modelling, one of the remaining challenges is the consistency of the proposed modelling approaches with the drivers' underlying behavioural process of decision-making.

By adopting a random utility maximization framework and a two-stage modelling process with sampling of alternatives, the main objective of this thesis is to improve the understanding of drivers' route choice behaviour. To fulfill this main objective, more specific goals and contributions are defined for this thesis that are summarized below.

1.4.1 Space Hierarchy and the Role of Anchor Points

Objective: It is well established in the literature that individuals orient themselves based on distinguished features of the route, called anchor points, and follow a hierarchical planning strategy influenced by the hierarchical representation of space. However, route choice models have mostly neglected the effect of space hierarchy and anchor points on route choice decisions. The first objective of this thesis is to propose a route choice modelling framework in which the effect of anchor points is incorporated along with the effect of route level attributes.

Contribution: The effect of space hierarchy and anchor points on drivers is captured using an "anchor-based nested" modelling approach, to promote the behavioural aspect of route choice models by considering both anchor points and route-level attributes. First, a Nested Logit (NL) structure is adopted, in which upper nests correspond to anchor points and lower nests include route alternatives. Second, a nested Logit Kernel (LK) model is estimated to capture the reciprocal effect of route level attributes and anchor points on route selection. Moreover, in the former model, the nested structure captures the shared unobserved components of the utility function among routes crossing the same anchor points, while in the latter, the adopted factor analytic approach accounts for the interdependencies and latent similarities.

1.4.2 Behavioural Classification

Objective: The classification of travellers' behaviours (such as pedestrians, bike users, bike-sharing members, car-sharing members, etc.) has been extensively studied in the literature. Although previous studies have shown that different categories of road users are observable, an explicit classification of car drivers route choice behaviours, based on their actual route choices, is missing from the literature. This research gap defines our next objective and contribution.

Contribution: This study sheds some light on taxi drivers' different route choice behaviours based on their revealed route choices. Taxi drivers' actual route choices are studied and classification methods are adopted to characterize their different route choice behaviours.

1.4.3 Specialized Data Collection

Objective: Route choice studies are rarely based on data collection methods tailored for route choice modelling purposes. Existing data collection methods mostly ignore the behavioural and attitudinal factors affecting individuals' responses. The increasing application of advanced choice models, reflecting the stochasticity of individuals' preferences and the complex nature of human decision-making behaviour, requires enhanced data collection methods collecting detailed data without significantly increasing respondent burden. This research aims at designing and implementing such a survey framework.

Contribution: A web-based survey is designed to provide a rich dataset, based on which reliable behavioural route choice models can be produced. The survey is designed to reveal drivers' consideration set of route alternatives, their final choices, and the latent behavioural traits that affect the formation of drivers' consideration set and influence their final choices. In short, it enables the analyst to investigate more closely some major challenges facing route choice modelling, such as the definition of route alternatives and how they are perceived by drivers, the characteristics of a considered set of route alternatives, and the role of different attributes (observable and latent) in route choice decisions.

1.4.4 Behavioural Traits and Latent Heterogeneity

Objective: It has been well-established in the literature that latent traits such as attitudes, perceptions, and lifestyle preferences affect individuals' decisions. Moreover, different segments of the population, might also have different choice behaviours. These latent factors and taste heterogeneity have mostly been ignored in previous route choice studies. Another objective of this research is to compare the route choice behavior of frequent versus occasional drivers and to provide a comprehensive framework to explicitly incorporate the effect of latent behavioural constructs as well as population segments' taste heterogeneity into the choice modelling process.

Contribution: To compare route choice behavior of frequent versus occasional drivers and to incorporate both the effects of segment heterogeneity and latent variables in their decision-making process, the LC-ICLV modelling framework is proposed by considering an Integrated Choice and Latent Variable (ICLV) model as the choice component of a Latent Class (LC) model. The ICLV model is used to bring in the role of the underlying behavioural constructs, while the LC component accounts for the taste heterogeneity across population segments.

1.4.5 Consideration Set of Route Alternatives

Objective: The way individual drivers derive their actual consideration set of route alternatives, and factors affecting the size and composition of these choice sets is a complex subject. Individuals' choice sets are dependent on objective constraints, such as route attributes as well as subjective criteria, such as individuals' attitudes, perceptions and experiences. The final objective of this thesis is to elaborate on this aspect of route choice modelling by observing drivers' actual consideration sets and analysing factors affecting their sizes.

Contribution: The relationship of six broad categories of factors on the size of the declared consideration sets is studied. Accordingly, factors having an incidence on the size of the considered choice sets are identified to illustrate that different sizes of considered choice sets can be associated to different types of choice behaviour.

1.5 Thesis Structure

This thesis is presented in 10 chapters and includes the contributions in the form of articles presented in five chapters (Chapter 4 to Chapter 8).

- This chapter (Chapter 1) presents a concise background on route choice modelling together with motivations, research objectives and contributions of this research.
- Chapter 2 presents a review of relevant research efforts, focusing on route choice modelling approaches, data collection methods, and factors affecting route choice decisions.
- Chapter 3 focuses on the general methodology and the scope of the presented work.
- Chapter 4 presents the article entitled "On the Role of Bridges as Anchor Points in Route Choice Modelling", published in "Transportation" journal.
- Chapter 5 presents the article entitled "Classifying Behavioural Dynamics of Taxi Drivers Operating Strategies Using Longitudinal Route Choice Data", which has been submitted for publication in the journal "Transportation Research Part F: Traffic Psychology and Behaviour".
- Chapter 6 presents the article entitled "An Online Survey to Enhance the Understanding of Car Drivers Route Choices", which has been presented at the 11th International Conference

on Transport Survey Methods (ISCTSC) and has been submitted for publication in the "Transportation Research Procedia" of the conference.

- Chapter 7 presents the article entitled "Frequent Versus Occasional Drivers: A Hybrid Route Choice Model", which has been submitted for publication in "Transportation Research Part F: Traffic Psychology and Behaviour" journal.
- Chapter 8 presents the article entitled "Factors Affecting Drivers' Consideration Set of Route Alternatives", which has been presented at Transportation Research Board 2018 Annual Meeting.
- Chapter 9 provides a general discussion on the main findings of this research.
- Chapter 10 summarises the most important contributions of this thesis and discusses the major conclusions, limitations and directions for future research in this area.

CHAPTER 2 MODELLING ROUTE CHOICE DECISIONS: A REVIEW

This section covers the state of the art in modelling route choice decisions in a discrete choice modelling framework. First, the discrete choice modelling framework is introduced, and approaches with restricted and unrestricted choice sets are described. The second part of this section presents a brief review of data collection methods adopted in route choice studies. Third, several types of attributes and factors affecting drivers' route choice decisions are discussed.

2.1 Route Choice Modelling: A Discrete Choice Framework

What is random utility discrete choice modelling? It is a modelling framework that stems from the consumer theory of microeconomics and is applied to understand and predict decision makers' choices. In this framework, observed choices are considered as discrete events and individuals' preferences are represented by a vector called *"utility"*, which captures the effect of different factors on the final choices. Also, individuals are viewed as rational decision makers who intend to maximize their perceived utilities. The utility function consists of two components presented in Equation 2.1.

$$U_{in} = V_{in} + \varepsilon_{in} \tag{2.1}$$

The deterministic part of the random utility function V_{in}, depends on the observed characteristics of each decision maker n. In order to take into account the stochasticity and imprecision caused by the lack of information, uncertainty and behavioural randomness of the decision maker, an error component is added to the utility vector (Ben-Akiva & Bierlaire, 2003).

Since every decision maker seeks to maximize his perceived utility, the utility of each alternative $j \in C_n$ is evaluated and the probability that a given alternative i is chosen by decision maker n from the feasible choice set C_n is given by Equation 2.2.

$$P(i|C_n) = P\left[U_{in} = \max_{j \in C_n} U_{jn}\right] \tag{2.2}$$

Various types of models have been developed based on specific assumptions on the structure of the deterministic component V_{in} or depending on the statistical distribution of the stochastic component ε_{in} of the utility function. The Multinomial Logit Model (MNL) is the simplest example of these models, which assumes that the error component of the random utility function is independently and identically distributed (i.i.d.) with a Gumbel distribution. The MNL formulation is illustrated in Equation 2.3,

$$P(i|C_n) = \frac{e^{V_i}}{\sum_{j \in C_n} e^{V_j}} \tag{2.3}$$

where $P(i|C_n)$ is the probability of choosing alternative i from C_n, C_n is the choice set of feasible routes for individual n, and V_i and V_j are the deterministic components of the utility functions for paths i and j, respectively. To model route choice decisions, two major schools of thought exist within the realm of discrete choice modelling: i) the two-stage choice modelling approach with restricted choice set, and ii) the recursive logit with unrestricted choice set. These approaches are discussed subsequently.

2.1.1 Two Stage Choice Modelling with Restricted Choice Set

Analysing choices using this approach requires knowledge regarding the set of alternative choices available to the decision maker from which the final decision has been made. This set is usually called the considered choice set (or consideration set) and contains a finite number of alternatives (Ben-Akiva & Bierlaire, 2003). In other words, in a two-stage choice process, decision makers reduce the total number of possible alternatives U (i.e. the universal choice set), to a smaller set of feasible alternatives, called the considered choice set C_n, in the first stage. Then, in the second stage they compare every alternative in the latter to find the one that maximizes their perceived utility (Ben-Akiva & Boccara, 1995; Bovy, 2009; Manski, 1977). The two stage modelling process can be formulated as

$$P(i) = \sum_{C_n \subset H_n} P(i|C_n)P(C_n) \tag{2.4}$$

where $P(i)$ denotes the probability of route i being chosen, $P(i|C_n)$ is the probability of choosing route i given C_n (the considered choice set of route alternatives for individual n), $P(C_n)$ denotes the probability of choosing the considered set C_n among all the possible non-empty subsets of route alternatives H_n. However, defining a probabilistic choice set of route alternatives (i.e. $P(C_n)$) is a very complex and mathematically intractable problem due to the size of H_n, and has never been used in real-world route choice applications.

However, the MNL model can be consistently estimated on a subset of alternatives using the approach discussed by McFadden (1978), in which the probability of choosing route i is conditional on the defined consideration set C_n, and an alternative specific correction term is added to correct for sampling bias. This conditional probability is calculated as:

$$P(i|C_n) = \frac{e^{\mu V_{in} + \ln q(C_n|i)}}{\sum_{j \in C_n} e^{\mu V_{jn} + \ln q(C_n|j)}} \tag{2.5}$$

where $\ln q(C_n|i)$ is the alternative specific correction term based on the probability of sampling C_n given that route i has been chosen.

The complexity of modelling route choice behaviour in a two-stage modelling approach arises from two main challenges: 1) the large number of possible routes connecting a given origin destination pair in a real world network, and 2) the complex correlation structure among the overlapping route alternatives. Choice set generation methods and advanced choice models have been developed to deal with these challenges. An overview of these methods is presented next.

2.1.1.1 Choice Set Generation

In route choice modelling, considered sets of alternatives are mostly latent to the analyst and are rarely observed (Hoogendoorn-Lanser, van Ness, & Bovy, 2005; Prato et al., 2012). In real world road networks, it is computationally very expensive and operationally not feasible to identify the universal choice set, i.e. to enumerate all the paths connecting a given OD pair. Hence, deterministic and stochastic route generation techniques have been adopted to create a subset of alternatives known as the master set, which approximates all the routes that are supposed to be known to the decision maker. However, this set may still be very large and may contain implausible and unattractive alternatives. Moreover, drivers are assumed to be aware of all of them and tirelessly compare their attributes to choose the best one. Due to the limited information processing abilities of drivers, spatial and temporal restrictions, and latent traits, such as attitudes and preferences, a set of spatiotemporal constraints and screening rules are usually adopted to delimit the consideration set, which is supposed to represent the actual set of routes from which drivers pick their final one (Ben-Akiva & Boccara, 1995; Bovy, 2009; Prato et al., 2012).

Since random sampling of alternatives in large universal choice sets is not efficient in terms of providing information, an importance sampling method is deemed to be more convenient and favorable to sample more attractive alternatives (Hess & Daly, 2010). Several deterministic and stochastic path generation methods, mostly based on repeated shortest path algorithms, have been proposed in the literature to form consideration sets.

Among the deterministic methods are k-shortest path (Papinski & Scott, 2011), link labelling (Ben-Akiva, Bergman, Daly, & Ramaswamy, 1984), link elimination (Azevedo, Costa, Madeira, & Martins, 1993), and link penalty (de la Barra, Perez, & Anez, 1993) methods. The k-shortest path algorithm is a generalization of the shortest path algorithm (Dijkstra, 1959), which finds the first k shortest paths connecting two points, based on a specific cost function. In route labelling, the analyst generates shortest paths based on different cost functions (i.e. labels), such as scenery, travel time, travel distance, number of lights, etc. The link elimination approach finds the next best route, based on a generalized cost function, by eliminating one or more links from the previously found shortest path. The elimination can be performed randomly or controlled by predefined criteria (Dhaker, 2012; Prato & Bekhor, 2007b; Rieser-Schüssler, Balmer, & Axhausen, 2013). Link penalty is an iterative approach in which the generalized cost function is gradually increased, and a new shortest path is calculated in each step. The iteration is repeated until a predefined number of route alternatives are generated (Dhaker, 2012; Rieser-Schüssler et al., 2013).

Stochastic route generation methods, however, randomly draw link impedances from probability distributions to account for individuals' preferences and perception errors. Then, repeated shortest path algorithms are used to select route alternatives based on these random generalized costs (Dhaker, 2012; Frejinger, 2008; Prato, 2009b).

The major downside of using these deterministic and stochastic methods is that they do not provide researchers with sampling probabilities of the selected alternatives. Model estimates based on these path generation techniques are biased, unless the sampling probability of every alternative in the universal set is equal, which is not the case in route choice modelling (Frejinger et al., 2009; Prato, 2009a). Several alternative approaches have been proposed to deal with this challenge. For instance, Manski (1977) proposed a full probabilistic method in which an inclusion probability (or sampling probability) is calculated for every possible route alternative. This approach is impractical and unmanageable in real world networks where the total number of possible route alternatives is very large (Prato, 2009b). Cascetta, Russo, Viola, and Vitetta (2002) adopted a two-stage process, where in the first step, a complete set of alternatives is generated for all the observed trips by maximizing a coverage factor between the generated set and the set of routes perceived as available. Then, in the second step, a binomial logit model is adopted to estimate the probability of including a given route in users' consideration sets. Frejinger et al. (2009) applied a biased random walk to sample a subset of paths and derived a sampling correction factor to estimate unbiased parameters. More recently, Flötteröd and Bierlaire (2013) used a Metropolis-Hastings (MH) algorithm to generate sample sets based on an arbitrary distribution providing the sampling probability of each alternative. This algorithm requires a road network and a definition of path weight as inputs. It uses an underlying Markov Chain process to sample alternatives and calculates its sampling probability without the need of normalizing it over the full choice set.

For more details and in-depth comparisons of the discussed methods the reader is referred to (Bovy, 2009; Frejinger, 2008; Frejinger & Bierlaire, 2010; Prato, 2009b; Rieser-Schüssler et al., 2013).

2.1.1.2 Route Choice Models

In the MNL model, which is a powerful tool for modelling choice behaviour, the likelihood of choosing alternative $A1$ over another alternative $A2$ is independent from whether a third alternative $A3$ is present. This feature, called the Independence of Irrelevant Alternatives (IIA), may not hold in real world route choice situations, where different alternatives share similar links. Hence, it may overestimate the choice probability of routes with common links.

Therefore, different approaches have been proposed to consider the correlations between route alternatives. The first approach is to maintain the logit structure and modify the deterministic part of the utility function. Models such as C-Logit (Cascetta, Nuzzolo, Russo, & Vitetta, 1996b), Path-Size Logit (Ben-Akiva & Bierlaire, 1999a) and Path-Size Correction Logit (Bovy, Bekhor, & Prato, 2008) models are in this category. The second modelling approach accounts for similarities in the stochastic part of the utility function and presents a closed-form expression for the choice probability. These models, also known as Multivariate Extreme Value (MEV) models, include Paired Combinatorial Logit, Cross Nested Logit and Generalized Nested Logit models. The third approach, i.e. the Non-MEV modelling approach, accounts for similarities in the stochastic part of the utility function and does not have a closed-form expression for the choice probability. This category of model includes Multinomial Probit, Mixed Logit, and Logit Kernel models (Dhaker, 2012; Prato, 2009a). A brief overview of these approaches follows.

2.1.1.2.1 Modifying the Deterministic Component

This approach maintains the simple structure of logit and adds a correction term to the utility function to consider the amount of similarity (or dissimilarity) between alternatives. This model takes the general form of:

$$P(i|C_n) = \frac{e^{V_i+\beta_{CF} * CF_i}}{\sum_{j\in C_n} e^{V_j+\beta_{CF} * CF_j}} \tag{2.6}$$

where CF_i and CF_j are the correction factor associated with path i and j, respectively, and β_{CF} is a parameter to be estimated. The correction factor is usually designed to increase the probability of choosing an independent route over choosing an alternative with overlap. Hence, β_{CF} is usually a negative value and is expected to reduce the utility of paths with common links.

Different types of correction factors have been proposed in the literature. For instance, Cascetta, Nuzzolo, Russo, and Vitetta (1996a) proposed a correction factor based on the amount of overlap and similarity of a route with other route alternatives:

$$C_i = \gamma_0 ln \sum_{j\in C_n} \left(\frac{d_{ij}}{\sqrt{d_i.d_j}}\right)^{\gamma_1} \tag{2.7}$$

where C_i is the correction factor known as commonality factor of path i, d_{ij} is the common length between path i and j, d_i and d_j are the lengths of paths i and j, respectively, and γ_0 and γ_1 are parameters to be estimated. Different formulations have been proposed for the commonality factor which can be found in (Cascetta et al., 1996a; Prato, 2009a; Ramming, 2002), however, the form of commonality factor to use is not clearly specified.

Ben-Akiva and Bierlaire (1999b) proposed a correction factor named Path-Size factor, and proposed the following formulation:

$$PS_i = \sum_{a\in\Gamma_i} \left(\frac{l_a}{L_i}\right) \frac{1}{\sum_{j\in C_n} \delta_{aj} \frac{L^*_{C_n}}{L_j}} \tag{2.8}$$

where, PS_i is the Path-Size factor for route i, Γ_i is the set of links in route i, l_a is the length of link a, L_i and L_j are the length of path i and j, respectively, δ_{aj} is the link-path incident variable (1 if link a is on path j; 0 otherwise), and $L^*_{C_n}$ is the length of the shortest path in the choice set. The first term in these equations correspond to the fraction of impedance imposed from a specific link, and the second term reflects the number of paths using the link. The generalized Path-Size factor also known as exponential Path-Size factor proposed by Ramming (2002) tends to decrease the influence of longer paths on the utility of shortest ones:

$$PS_i = \sum_{a\in\Gamma_i} \left(\frac{l_a}{L_i}\right) \frac{1}{\sum_{j\in C_n} \left(\frac{L_i}{L_j}\right)^{\delta_{PS}} \delta_{aj}} \tag{2.9}$$

where, δ_{PS} is a parameter to be estimated. Bovy et al. (2008) proposed a revised version of the Path-Size factor, called the Path-Size Correction (PSC) factor that takes the following form:

$$PSC_i = -\sum_{a\in\Gamma_i} \left(\frac{l_a}{L_i} ln \sum_{j\in C_n} \delta_{aj}\right) \tag{2.10}$$

Frejinger et al. (2009) extended the Path-Size factor to take into account the correlation of each alternative with all the possible paths in the true (i.e. universal) choice set. The Extended Path-Size (EPS) factor is defined as:

$$EPS_{in} = \sum_{a\in\Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j\epsilon\varphi_n} \delta_{aj}\omega_{jn}} \tag{2.11}$$

where, $\omega_{jn} = \frac{k_{jn}}{q_{jn}}$ is an extension factor with a value equal to 1 if $\delta_{aj} = 1$ or $q_{jn}R_n \geq 1$, and $1/(q_{jn}R_n)$ otherwise; where R_n denotes the total number of paths drawn with replacement from the universal choice set, q_{jn} is the sampling probability of path j, and k_{jn} is the empirical frequency or the actual number of times path j is drawn with replacement from the universal choice set, by the choice set generation algorithm.

2.1.1.2.2 Modifying the Stochastic Component

These models account for the correlation between route alternatives by modifying the stochastic component of the utility function. Multivariate Extreme Value (MEV) models account for these similarities by maintaining a closed form (e.g. Nested Logit, Cross-Nested Logit, etc.). For instance, in the cross-nested logit structure, applied by Vovsha and Bekhor (1998) in a route choice context and later adopted by Lai and Bierlaire (2015), the correlation between different route alternatives is taken into account by modifying the modelling framework, so that each link of the network constitutes a nest and each route belongs to several nests. Also, to consider taste variation or correlation between unobserved factors, non-MEV model structures (such as Multinomial Probit, Logit Kernel, etc.) are adopted.

In this section, we only present the Nested Logit, Cross Nested Logit and the Logit Kernel modelling frameworks, which are also relevant to subsequent sections. For more details regarding these models, the enthusiastic reader is referred to (Dhaker, 2012; Frejinger, 2008; Prato, 2009b).

2.1.1.2.2.1 Nested Logit

The Nested Logit formulation, was proposed by Ben-Akiva (1973) and proved to be consistent with the stochastic utility maximization theory by McFadden (1978). It is an extension of the Multinomial Logit and captures some of the unobserved similarities among alternatives by dividing the choice set into several nests. These nests are considered to be collectively exhaustive and mutually exclusive in covering the considered alternatives. Every nest contains a subset of alternatives sharing a particular characteristic, independent from other subsets of alternatives in other nests. In other words, the probability of choosing an alternative from a nest is considered to be independent from alternatives in other nests. The probability of choosing an alternative can be expressed as the product of the conditional probability of choosing that alternative given a particular nest and the choice probability of that respective nest (Ben-Akiva, 1973; Guevara & Ben-Akiva, 2013). Within this framework, the probability of choosing alternative i by individual n given the true (i.e. universal) choice set C_n is

$$P(i|C_n) = \frac{e^{V_{in}} G_i(e^{V_{1n}}, \dots, e^{V_{J_n}})}{G(e^{V_{1n}}, \dots, e^{V_{J_n}})} \tag{2.12}$$

$$G_i(e^{V_{1n}}, \dots, e^{V_{J_n}}) = \frac{\partial G}{\partial e^{V_{in}}}(e^{V_{1n}}, \dots, e^{V_{J_n}}) \tag{2.13}$$

where G is a non-negative differentiable MEV generating function, G_i is its partial derivative with respect to $e^{V_{in}}$, V_{in} specifies the systematic part of the utility function, and J_n is the number of alternatives in C_n. The probability of choosing alternative i from the true choice set can be written as:

$$P(i|C_n) = \frac{e^{V_{in} + \ln G_i(e^{V_{1n}}, \dots, e^{V_{J_n}})}}{\sum_{j=1}^{J_n} e^{V_{jn} + \ln G_j(e^{V_{1n}}, \dots, e^{V_{J_n}})}} \tag{2.14}$$

The partial derivative of the MEV generating function for Nested Logit G_i given the true choice set C is calculated using:

$$G_i(C_n) = G_i(e^{V_{1n}}, \dots, e^{V_{Jn}}) = \mu e^{V_{in}(\mu_m - 1)} \left(\sum_{i=1}^{J_m} e^{\mu_m V_i} \right)^{\frac{\mu}{\mu_m} - 1} \tag{2.15}$$

where μ and μ_m are scale parameters for the model and its nests, respectively, where $\mu/\mu_m \leq 1$, m is the nest including alternative i, and J_m is the number of alternatives in nest m. The Nested Logit modelling framework and its application in route choice modelling are presented in more detail in Chapter 4.

2.1.1.2.2.2 Cross Nested Logit

The Cross Nested Logit (CNL) model, also referred to as Link Nested Logit, was first applied in the context of route choice by Vovsha and Bekhor (1998). In this model, contrary to the Nested Logit model, alternatives can belong to several nests and the overlapping problem of route alternatives is handled by nesting parameters. Also, the upper level nests are formed by common links and routes from the lower level nests. The probability of alternative i being chosen from the choice set C_n and a total of M nests is given by

$$P(i|C_n) = \sum_{m=1}^{M} P(C_{mn}|C_n) P_n(i|C_{mn}) \tag{2.16}$$

where $P(i|C_{mn})$ is the conditional probability of alternative i being chosen in the nest m, given by

$$P(i|C_{mn}) = \frac{\alpha_{mi} e^{V_{in}}}{\sum_{j \in C_{nm}} \alpha_{mj} e^{V_{jn}}} \tag{2.17}$$

where α_{mi} is the inclusion weight for each alternative i in nest m with a value between zero and one, representing the degree to which alternative i is a member of nest m (Ramming, 2002; Vovsha & Bekhor, 1998). This coefficient captures the similarity within a nest and has the following formulation (Bekhor & Prashker, 2001):

$$\alpha_{ai} = \frac{l_a}{L_i} \delta_{ai} \tag{2.18}$$

where l_a and L_i are the length of link a and route i, respectively, and δ_{ai} is the link-route incident variable which is considered to be 1 when link a is part of route i. The marginal probability of the nest C_{mn} being chosen is

$$P(C_{mn}|C_n) = \frac{e^{V_{C_{mn}}+\mu_m I_{C_{mn}}}}{\sum_{l=1}^{M} e^{V_{C_{ln}}+\mu_m I_{C_{ln}}}} \quad (2.19)$$

where μ_m is the nesting coefficient which captures the similarity between nests and is considered constant by Vovsha and Bekhor (1998). $I_{C_{mn}}$ takes the following form :

$$I_{C_{mn}} = ln \sum_{j \in C_{mn}} \left(\alpha_{mj} e^{V_{jn}}\right)^{1/\mu_m} \quad (2.20)$$

Replacing the three previous equations in Equation 2.16, the conditional probability of choosing alternative i from C_n is given by

$$P_n(i|C_n) = \frac{\sum_{m=1}^{M} (\alpha_{mi} e^{V_i})^{\frac{1}{\mu_m}} \left(\sum_{j \in C_{mn}} (\alpha_{mj} e^{V_i})^{\frac{1}{\mu_m}}\right)^{\mu_m - 1}}{\sum_{m=1}^{M} \left(\sum_{j \in C_{mn}} (\alpha_{mj} e^{V_i})^{\frac{1}{\mu_m}}\right)^{\mu_m}} \quad (2.21)$$

The application of this model for a real-world network can become very computationally expensive because of the nesting structure (Cascetta & Papola, 2001).

2.1.1.2.2.3 Logit Kernel (LK)

Logit Kernel, which is a combination of Probit and Logit models, was first proposed by Bolduc and Ben-Akiva (1991).The random component of its utility function is composed of a probit-like term, which captures the interdependencies among alternatives, and an i.i.d. Gumbel distributed random component. The interdependencies between alternatives can be explicitly specified using a factor analytic approach, proposed by McFadden (1984). This approach accommodates different error structures and reduces the estimation complexity of the model (Bekhor, Ben-Akiva, & Ramming, 2002; Bierlaire & Frejinger, 2005). The utility function for individual n is defined by:

$$U_n = X_n \beta + F_n T \zeta_n + \nu_n \quad (2.22)$$

where U_n is the utility vector of size $(J_n \times 1)$, and J_n is the number of alternative in the choice set C_n; X_n is the matrix of explanatory variables of size $(J_n \times K)$; β is the vector of unknown parameters of size $(K \times 1)$; F_n is the factor loading matrix of size $(J_n \times M)$; T is a diagonal matrix of the standard deviation of each factor of size $(M \times M)$; ζ_n is the vector of i.i.d. random variables with zero mean and unit variance of size $(M \times 1)$; and ν_n is the vector of i.i.d. Gumbel distributed random term with zero location of size $(J_n \times 1)$, a scale equal to μ, and a variance equal to $(\pi^2/6\mu^2)$. The LK model can replicate any error structure and approximate any random utility model (Ben-Akiva, Bolduc, & Walker, 2001; McFadden & Train, 2000; Walker, Ben-Akiva, & Bolduc, 2004). In a Nested Logit analog to the LK model, also known as the nested LK model, F_n is defined to be the alternative-nest incident matrix and is obtained by defining a dummy variable for each nest that equals 1 if an alternative belongs to that particular nest, and 0 otherwise. Moreover, ζ_n is usually assumed to be normally distributed $N(0,1)$, and T captures the amount of correlation between alternatives belonging to the same nest (Walker et al., 2004; Walker, 2001). If the factors ζ_n are known, the probability of choice i given ζ_n is estimated by:

$$\Lambda(i|\zeta_n) = \frac{e^{\mu(X_{in}\beta + F_{in}T\zeta_n)}}{\sum_{j=1}^{J_n} e^{\mu(X_{jn}\beta + F_{jn}T\zeta_n)}} \tag{2.23}$$

Since ζ_n is unknown, the unconditional probability takes the following form:

$$P(i) = \int_{\zeta} \Lambda(i|\zeta_n) \prod_{m=1}^{M} \phi(\zeta_m) d\zeta \tag{2.24}$$

where $\phi(\zeta_m)$ is the standard univariate normal density function, and $\prod_{m=1}^{M} \phi(\zeta_m)$ represents the joint density function of ζ. Since the probability function does not have a closed form, it is approximated through simulation:

$$\hat{P}(i) = \frac{1}{D} \sum_{d=1}^{D} \Lambda(i|\zeta_n^d) \tag{2.25}$$

where D is the number of simulation draws and ζ^d denotes draw d from the distribution of ζ. The application of a Logit Kernel model in route choice modelling is presented in more detail in Chapter 4.

2.1.2 Recursive Logit with Unrestricted Choice Set

Fosgerau, Frejinger, and Karlstrom (2013) presented a recursive logit model which can be consistently estimated in a real-world road network without the need of path sampling. The model relates a sequential link based route choice model and the finite MNL model, in a dynamic discrete choice modelling framework. Also, in order to account for the correlation in path utilities, a link additive correction factor, called Link Size, which corrects the utility function in a similar way to the Path-Size attribute, is added to the utility function.

In a further study, Mai, Fosgerau, and Frejinger (2015) extended the recursive logit model, allowing path utilities to be correlated in a way similar to the nested logit model, and therefore relaxed the IIA property associated to the MNL model by assuming that scale parameters are link specific. In a recent study, Mai (2016) formulated a generalized recursive logit model, called the Recursive Network MEV (RNMEV) model, which can approximate any additive random utility model (e.g. the Nested Logit, or the Cross Nested Logit). Moreover, a recent study by Mai, Bastin, and Frejinger (2016) used a decomposition method to reduce the number of linear systems to be solved and reduce the computation time required for model estimation.

For a more in-depth discussion on the formulation, estimation procedure and application results of the recursive logit model, the reader is referred to (Fosgerau et al., 2013; Mai, 2016; Mai et al., 2016; Mai, Bastin, & Frejinger, 2017; Mai et al., 2015; Zimmermann, Mai, & Frejinger, 2017a).

2.2 Route Choice Data Collection

In general, data collection methods for choice modelling in transportation studies can be classified into three broad categories, namely Revealed Preferences (RP) surveys, Stated Preferences (SP) surveys, and passive data. These methods are concisely discussed below.

2.2.1 Revealed Preferences (RP)

RP surveys require respondents to describe their actual chosen routes and may require additional information on factors affecting their decisions (Koller-Matschke, Belzner, & Glas, 2013; Parkany, Du, Aultman-Hall, & Gallagher, 2006b; Ramaekers, Reumers, Wets, & Cools, 2013). Since RP route choice surveys mostly focus on a single choice task, they may not provide sufficient information regarding drivers' choices and the relative importance of factors affecting them. Hence, they may not be very adequate for prediction purposes (Ortúzar & Willumsen, 2011).

2.2.2 Stated Preferences (SP)

SP surveys require respondents to choose between a series of hypothetical options based on their preferences and have been widely used in transportation modelling. In SP route choice surveys, respondents are usually asked to choose between some hypothetical route alternatives, for which some details are provided. The major advantage of SP over RP surveys is that they can capture information on alternatives and attribute combinations that do not exist in real life. Moreover, SP surveys are the best way to collect responses for policies which are not yet implemented and do not exist. SP questions can be asked in a way to reveal the preferences of respondents among several hypothetical alternatives with different combinations of attributes, either by asking them to choose exactly one of the alternatives, or to rank or rate them (Hensher, 1994). Table 2.1 compares few characteristics of SP and RP surveys (Morikawa, 1989; Sanko, 2002).

Table 2.1: Comparison of RP and SP surveys

	RP data	**SP data**
Preference information	Result of actual behaviour	Hypothetical scenarios
	Consistent with the behaviour in real world	Possibility of inconsistency with the behaviour in real world
	Choice results	"Ranking", "Rating", "Choice", and etc.
	Actual chosen alternative	Hypothetically chosen alternatives
Evaluated attributes	Measurement error	No measurement error
	Limited range of attributes' levels	Wider range of attributes' levels
	Uncontrolled collinearity among attributes	Controlled collinearity among attributes
Choice set	Unknown	Known
Choice scenario	Usually one per respondents	More than one per respondents
Consistency with actual behaviour	More	Less

Questions that ask to rank and/or rate alternatives are known as *Conjoint Analysis* questions. *Rank-order* questions ask respondents to order alternatives based on their preferences. Rank-order data can be translated into *choice* data by defining each rank as the chosen alternative between a choice set including its lower alternatives. It is stated that rank-order data produces less reliable information for lower ranks. *Rating* questions highly increase respondents' burden, since they have to provide both the order and the degree of preference for each question. In this type of questions, respondents are asked to rate different alternatives according to a pre-defined rating scale. *Stated*

Choice responses, which are identical to first-order ranking responses, are used to identify the ultimate choice of a respondent considering all the alternatives (Hensher, 1994). In *Contingent Valuation* technique, which is principally used to assess the Willingness to Pay (WTP), respondents have to specify the amount that they are willing to pay for a specific alternative or policy. This method can provide valuable information on different aspects of a proposed policy in terms of respondents' WTP (Ortúzar & Willumsen, 2011).

Although SP data is not the best way to determine demand levels, it can be compared to revealed choices and can be used to complement RP data and improve the capability of a prediction model. Choice data can be directly used for prediction purposes using discrete choice models. Several studies have presented theoretical frameworks to combine RP and SP data so that they can be used together to complement each other (Brownstone, Bunch, & Train, 2000; Mark & Swait, 2008; Morikawa, 1989; Phaneuf, Taylor, & Braden, 2013; Sanko, 2002).

2.2.3 Passive data

Passive data is the third data collection method adopted for route choice modelling purposes and includes technologies such as GPS, automatic plate recognition, Bluetooth, and Wi-Fi. In recent years, the prevalent use of GPS technology has provided researchers with an abundance of high-resolution geospatial data. GPS data can be collected using GPS devices or smartphones and has been used in several route choice studies (Jan, Horowitz, & Peng, 2000; Li, 2004; Li, Guensler, & Ogle, 2005; Papinski, 2010; Papinski & Scott, 2011; Papinski, Scott, & Doherty, 2009; Ramaekers et al., 2013). Although GPS traces can also be classified as revealed preferences data, their inclusion in the passive data category is mainly due to the different nature of their collection, which does not necessarily require a respondent to fill a questionnaire.

One of the major contributions of GPS data in transportation planning may be its potential to complement (and eventually replace) travel surveys. Traditionally, travel surveys use a paper-based or computer-based interface to collect daily travel diaries. Several problems are associated with travel diaries such as trip underreporting, respondent fatigue, and inaccuracies regarding the time frame and location of reported trips.

Several studies have shown that a considerable amount of trips are underreported in travel diaries and that these missing trips are not evenly distributed among the population (McGowen, 2006).

Respondents may not remember all the trips they have made or may not be willing to report some of them, or some trips (such as short trips) may not seem important to report. Some respondents may not be willing to participate in the survey because it is too long. Also, the quality of data collected over an extended period may decrease in each successive day, because the respondents are fatigued and put less effort in completing the survey. The burden of taking long and detailed surveys discourages public participation. Moreover, respondents may have inaccurate perceptions of time, speed, or delay, which may introduce uncertainty into the data set. For instance, a respondent may have a longer perception of his/her waiting time compared to his/her in-vehicle time (Bricka, Zmud, Wolf, & Freedman, 2009; Bricka, Sen, Paleti, & Bhat, 2012; Wolf, 2004). GPS data record all trips, reduce respondents' burden and collect precise coordinates and timestamp information. Using GPS datasets in parallel with travel surveys can provide valuable information regarding drivers' route choice behaviour (Jan et al., 2000; Li, 2004; Papinski, 2010).

Despite all the advantages of using GPS devices, there are numerous difficulties associated to the collection and treatment of GPS data as well. For instance, one can forget to recharge the device so the battery will fail during the experiment and the complete trajectory will not be available; one can forget the device at home; signal loss occurs inside tunnels and urban canyons; it may take some time for the device to acquire the position after a cold start; data may be lost during the transfer procedure; or data accuracy may be low. Another limitation of GPS, mostly concerning route choice studies, is that traffic conditions on other possible route alternatives is unknown and that the considered choice sets, from which individuals make their final choices, remain unidentified (Tawfik, Rakha, & Miller, 2010a). Readers are referred to (Chen, 2013a; Frejinger, 2008; McGowen, 2006; Ortúzar & Willumsen, 2011) for more details regarding GPS travel surveys.

2.2.4 Survey Design

In this section, we concisely discuss the steps that need to be taken into consideration while designing a survey. Designing a survey starts with defining the objectives of the survey and the information that needs to be acquired. A clear definition of the target question is imperative and may be the most important step of the design. Defining an appropriate analytical method, which depends on the type and aggregation level of the answer, is equally important.

Once the objective is clearly stated, attributes that are considered to play an important role in the decision-making process should be identified. For each attribute, we need to determine the type of answer that we seek, in order to select a proper form for its corresponding question. Questions should be clear, comprehensible and as simple as possible. In SP surveys, respondents are usually asked to choose, rank or rate different hypothetical alternatives according to their preferences. Rating questions are probably the most demanding type of questions and choice questions are the easiest ones to interpret for prediction purposes (Hensher, 1994; Sanko, 2002).

To create the hypothetical scenarios, analysts have first to specify different levels for each attribute. The considered levels should be realistic and believable. Furthermore, their total number should be reasonable and pragmatic for the design of the survey. The number of considered levels for a specific attribute is related to its importance in the study, linear or non-linear effect on the outcome, and its interaction with other attributes (Hensher, 1994; Sanko, 2002). Different statistical methods are then used to combine different levels of attributes and create hypothetical scenarios, namely orthogonal designs (such as full factorial and fractional factorial designs), and efficient designs (such as Bayesian efficient design). Full factorial designs are used where all the possible combinations of different attributes are considered in the creation of scenarios. Since it exhaustively includes every possible combination, the total number of scenarios can get very large. Thus, fractional factorial designs are used to reduce the number of scenarios. Orthogonal designs mainly aim at minimizing the correlation in the data for estimation purposes. Orthogonality means that there is no collinearity between main attributes as well as their products. In other words, having attributes A and B, there would be no correlation between A, B and A*B. However, efficient designs aim at minimizing the standard errors of parameter estimates (Hensher, 1994; Metrics, 2012; Sanko, 2002).

A further step in survey design is to determine a representative sample of the population. A sample can be defined as a subset of instances representing a larger population for which information is sought. The sample population can be chosen randomly from the entire population, i.e. simple random sampling, or from various strata of the population, i.e. stratified random sampling. Defining the sample size is not very straightforward and depends on various determinants such as the population size, the variability of the parameters in the population, the accuracy needed, as well as time and budget. In the case of internet-based surveys, it is also very important that the selected responses represent the actual population. The proper use of oversampling or under-sampling of some strata of the population can account for biases in the studied population (Ortúzar & Willumsen, 2011).

It is also important to choose the best interface for the survey to encourage participants. A survey can be directly completed by the respondent himself (self-administered questionnaires) or through an interviewer (personal interview surveys), and several mediums can be adopted such as paper, phone, computer, and internet (Hensher, 1994; Sanko, 2002).

Figure 2.1 illustrates a schema of the most important steps of survey design. For more detailed discussion on different strategies of data collection, sampling methods, and design of SP surveys, the reader is referred to (Hensher, 1994; Metrics, 2012; Ortúzar & Willumsen, 2011; Richardson, Ampt, & Meyburg, 1995; Rose & Bliemer, 2008; Sanko, 2002).

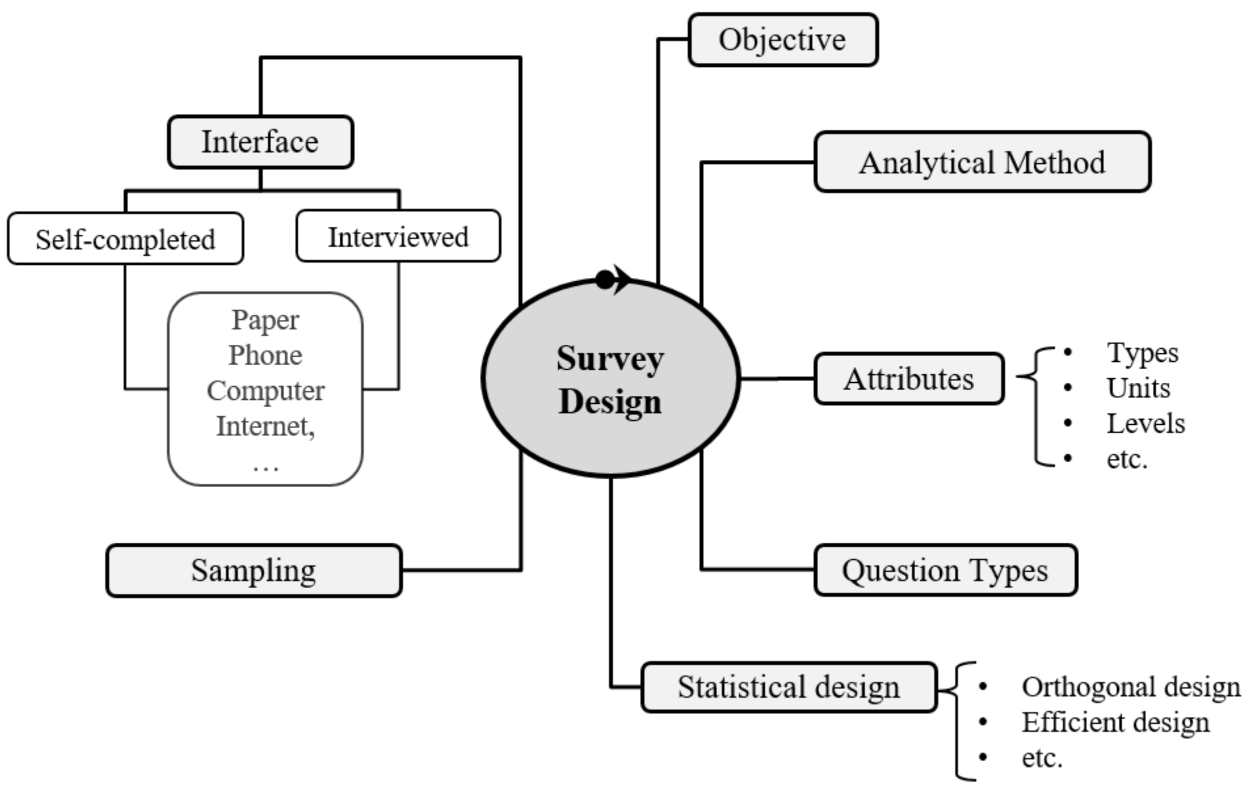


Figure 2.1: Steps in designing a Stated Preference (SP) survey.

2.3 Factors Affecting Drivers' Route Choice Decisions

According to the utility maximization framework, drivers tend to choose their paths between their origins and destinations in a way that maximizes their perceived utility. Jan et al. (2000) classified factors affecting route choice decisions into four main categories, illustrated in Table 2.2.

Table 2.2: Factors affecting route choice (Jan et al., 2000)

Traveller	Age, sex, life cycle, income level, education, household structure, race, profession, number of drivers in household, number of cars in household, etc.	
Route	Road	Travel time, travel cost, speed limits, waiting time. Type of road, width, length, number of lanes, angularity, intersections, bridges, slopes, etc.
	Traffic	Traffic density, congestion, number of turns, stop signs and traffic lights, travel speed, parking, probability of accident, reliability and variability in travel time, etc.
	Environment	Aesthetics, land use along route, scenery, easy pick-up/drop-off, etc.
Trip	Trip purpose, time budget, time of the departure, mode use, number of travellers	
Circumstances	Weather conditions, day/night, accident en route, route and traffic information, etc.	

However, these categories do not include geographical and behavioural factors, which are also found to significantly affect drivers' route choice decisions. Bovy and Stern (1990a) summarized factors influencing travel behaviour into four categories, illustrated in Figure 2.2.

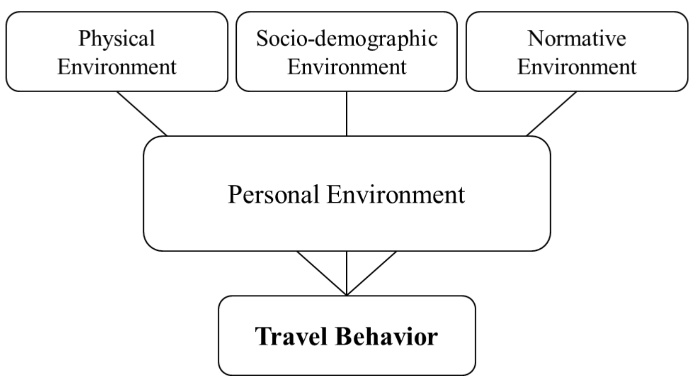


Figure 2.2: Factors affecting travel behaviour (Bovy & Stern, 1990a)

The physical environment component includes network characteristics and travel alternatives. The socio-demographic environment component is composed of factors such as household characteristics, age, gender, etc. The normative environment component encompasses the set of norms and values derived from the society. These three factors along with the personal environment, which incorporates the personality and attitudes of the traveller, affect the travel behaviour of the decision maker.

Another study by Bovy and Stern (1990a) presents a conceptual framework representing travellers' decision-making process (see Figure 2.3), in which the experimental and mental process of route choice, presented on the left hand side of the chart, influences the successive series of considered alternatives presented on the right hand side of the chart, leading to the final decision.

Moreover, attitudes and personal preferences are demonstrated to influence route choice decisions. These subjective characteristics can also be affected by the acquired knowledge of the road network, the level of information, and experience. Each route choice experiment can affect our personal preferences by its satisfactory or unsatisfactory outcome. It is argued that the repetition of satisfactory results can become a stable preference, and drivers tend to repeat the same choice forming a commute pattern (Ben-Akiva, De Palma, & Isam, 1991; Dia, 2002).

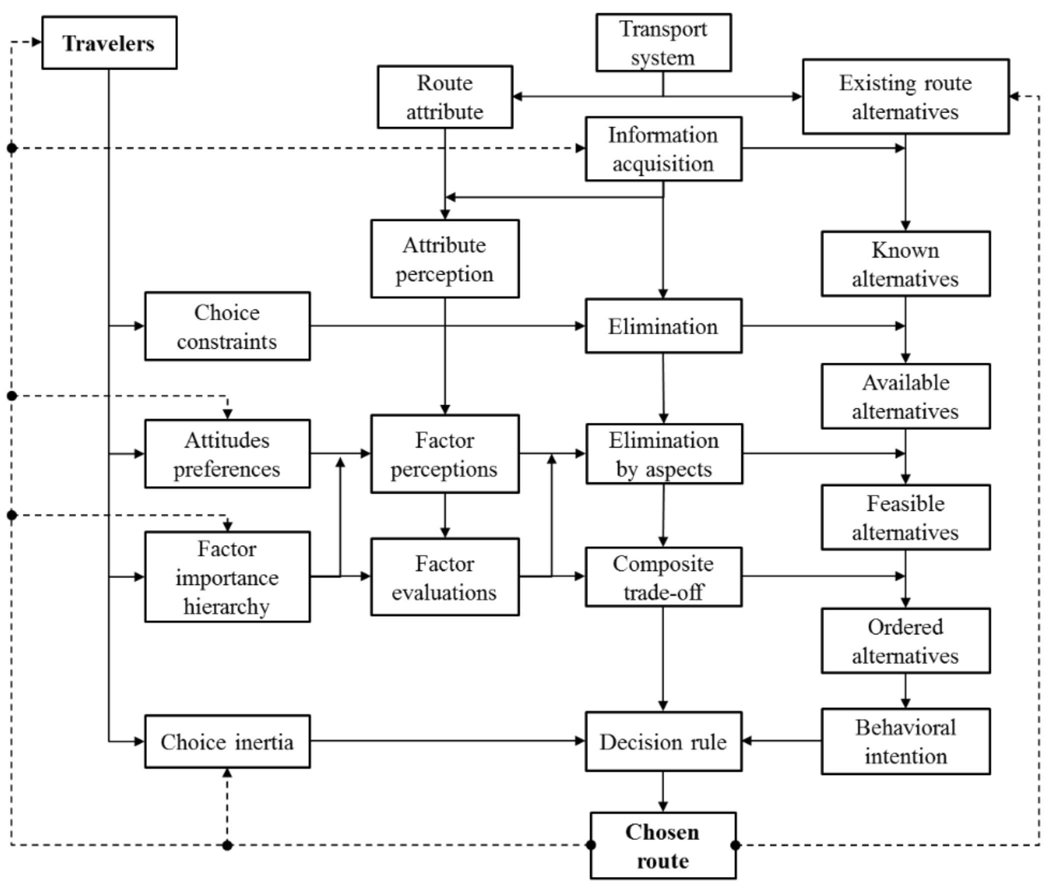


Figure 2.3: Conceptual framework of travellers' choice behaviour (Bovy & Stern, 1990a)

Many factors have been found to have an incidence on route choice decisions. Abdel-Aty and Jovanis (1997) argued that in addition to travel time, as the most important factor, travel time reliability, directness and congestion are among important factors influencing drivers' route choices. Peeta, Ramos, and Pasupathy (2000) collected stated preferences data on an expressway in the region of northwestern Indiana and showed that age and sex were the only socioeconomic characteristics that had a significant effect on travellers' route choices. Also, in a study by Cascetta et al. (2002), it was illustrated that network topology, socio-economic factors, and level-of-service significantly affect drivers' route choices.

Mannering (1989) investigated commuters' route choices and concluded that both traffic conditions and socio-economic characteristics play important roles in route choice decisions and the frequency of route changes. It has been found that unmarried people and younger commuters tend to change their routes more frequently than their married and older counterparts. Also, it has been stated that men change their routes more frequently. In a further study, Mannering, Kim, Barfield, and Ng (1994) investigated commuters' route changing behaviour and found that the frequency of route changes augments as the commute time becomes longer. It has been found that men are more likely to change route than women, and that individuals with high salary change their routes more frequently. Delay acceptance at work and familiarity with the road network were found to increase the frequency of route changes. Abdel-Aty (1994) found that high income, high level of education and high salary are among socio-demographic factors that increase the likelihood of using multiple routes, while commute distance was not found to be a significant factor. Also, men were found to change routes more frequently. Another study in Japan showed that more educated drivers are less likely to divert from their usual routes (Gan & Chen, 2013), while Jan et al. (2000) found that path deviation increases as travel distance increases. Parkany et al. (2006b) used GPS data for 125 drivers over a period of 10 days and concluded that men tend to change routes more frequently and are more likely to use freeways compared to women. Elderly people were found to have a lower tendency to change routes even when they become unexpectedly congested. Also, trip purpose was found to have a significant effect on route choice decisions, contrary to the number of signals, number of intersections and use of freeway.

Jou and Yeh (2013) studied the effect of toll rates on drivers' route choice behaviour. A stated preference survey via a questionnaire was conducted and a mixed logit model was developed to capture individuals' preference heterogeneity. Results illustrated that travel time, choice inertia, frequency of freeway use, time of freeway use, trip purpose, toll rate, personal income, and travel distance are among important factors affecting drivers' route choices.

Ramaekers et al. (2013) found that individuals prefer to drive on primary roads for work trips, in contrast to leisure trips. In addition, they found that peak and non-peak time periods do not have an influence on route choices. They also illustrated that beside travel time and travel distance, socio-demographic variables, personal income, and topography affect drivers' route choices. Cools, Moons, and Wets (2009) studied the effect of weather conditions on daily traffic volume and found that bad weather conditions reduce traffic volume, while high temperatures increase it. Papinski et al. (2009) explored the observed and planned route choices of thirty-one work commuters using GPS data and a questionnaire. They concluded that travel distance, stop signs, traffic lights, route directness, familiarity with the network and traffic conditions influence drivers' route choice behaviour.

Tawfik et al. (2010a) studied drivers' perceptions and experiences using a driving simulator as well as an initial and a final questionnaire. Results showed that drivers perceive travel speed better than travel time and these perceptions might affect their choices. Tawfik, Rakha, and Miller (2010b) found observable differences in the learning behaviour of drivers and categorized them based on their route choice patterns into four different categories. In a further study, Tawfik, Szarka, House, and Rakha (2011b) found that demographic characteristics and inertia can significantly affect route choice behaviour. The four learning patterns found in their previous studies were also found to have significant effects on route choice decisions and were used to develop disaggregated route choice models.

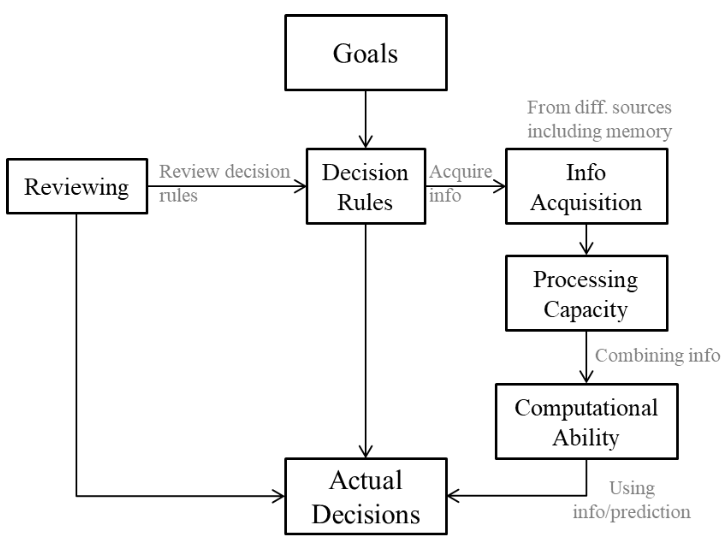


Figure 2.4 : Drivers' decision-making process (Ben-Akiva et al., 1991)

Another important factor affecting drivers' route choices is their level of information. Information can be acquired through direct observation, pre-trip or on-road information. According to Ben-Akiva et al. (1991), the fact that each driver chooses a different route can also be attributed to different levels of knowledge and information, different capacities to combine these information, and different computational and prediction abilities (see **Error! Reference source not found.**).

Route choice decisions are influenced by beliefs, which are in turn the outcomes of experiences. Experiences may be influenced by information and may in turn affect beliefs formed by previous experiences. Information available to drivers may be categorized as: 1) historical information, describing traffic conditions within previous time periods, 2) current information, describing the current state, and 3) predictive information, describing the expected traffic conditions during the travel period (Ben-Akiva et al., 1991). Since drivers are mostly affected by network conditions during their travel, the latter is the most profitable type of information, and yet the most difficult to obtain.

Several studies investigated the effect of information on route choice decisions and route changing behaviour of drivers. Maio, Vitetta, and Watling (2013) studied the effect of experience and day to day information update on route choice models. They argued that the inertia and habit of taking the same route prevents users from diverting to a better route when circumstances change. However, these habits may change over time because of perception updating due to information and experience acquired. They concluded that the learning process from experiments is influenced by the availability and type of information provided. Koller-Matschke et al. (2013) used GPS, interviews and driving simulator data, and showed that travel time is the main factor influencing route choice behaviour, and that travellers stay on the main road because of inertia, unless there is a considerable amount of time saving using other alternatives.

Ben-Elia and Shiftan (2010) investigated the effect of information and the effect of learning from experience via a mixed logit model. Socio-economic factors, route attributes, learning, and driving experience were among the factors considered in the modelling. Socio-economic characteristics were not found to have a significant effect and were excluded from the model. Results showed that providing information and the type of information provided affected the choice behaviour. Moreover, providing information has been found to be more effective on situations where drivers lack experience and are unfamiliar with the traffic conditions.

Iida, Uno, and Yamada (1994) investigated the relation between the mechanism and the quality of information transmitted and drivers' route choice behaviour, and found that the quality of information affects the willingness of drivers to use the information. Abdel-Aty (1994) investigated route changing behaviour based on the level of information. Results showed that individuals who listened to pre-trip traffic reports were more likely to change routes.

Peeta et al. (2000) investigated the relationship between the content of Variable Message Signs (VMS) and the frequency of route changes via an on-site SP survey and found that drivers who trust the provided information have a higher propensity to divert. A further study on the effect of Graphical Route Information Panel (GRIP) in the United States showed that information can have significant effects on drivers' route choices. Also, it has been demonstrated that driver's age, gender, and familiarity with the road network have significant influences on the level of knowledge acquired (Aitken, Conway, & Walton, 2012).

For more details regarding the role of information in route choice decisions the reader is referred to (Aitken et al., 2012; Ben-Elia, Erev, & Shiftan, 2008; Ben-Elia & Shiftan, 2010; Chatterjee, Hounsell, Firmin, & Bonsall, 2002; Gan & Chen, 2013; Gan, Bai, & Wei, 2013; Iida et al., 1994; Jou & Yeh, 2013; Koller-Matschke et al., 2013; Kusakabe, Sharyo, & Asakura, 2012; Levinson, 2003; Maio et al., 2013; Parvaneh, Arentze, & Timmermans, 2012; Peeta et al., 2000; Tian, Huang, & Liu, 2010; Wardman, Bonsall, & Shires, 1997; Yu & Peeta, 2011).

2.4 Synthesis

Among the various approaches that have been adopted to model drivers' route choice decisions, Random Utility Maximization (RUM) modelling framework has received considerable attention in previous studies. This section reviewed the main components of route choice modelling in a RUM modelling framework, namely the modelling approaches, the data collection methods, and important factors affecting route choice decisions. Table 2.3 summarizes some of the previous studies (in chronological order) and compares them based on the three main components of route choice modelling discussed above. Although the table does not encompass all the previous route choice studies, this list has been selected to provide a wide spectrum of research covering various data collection methods, considered factors, and modelling approaches.

Table 2.3: Summary of selected route choice studies

Study	Data Collection			Factors Considered						Modelling Approach	Significant Attributes *
	RP[a]	SP[b]	GPS	Obs[a]	CCS[b]	Demo[c]	Fact[d]	Behvr[e]	Percp[f]		
Iida et al. (1994)	☐	☒	☐	☐	☐	☐	☐	☐	☐	Linear models	28, 30
Abdel-Aty and Jovanis (1997)	☒	☒	☐	☒	☐	☐	☒	☐	☒	Binary logit	1, 2, 3, 4, 5, 6
Peeta et al. (2000)	☐	☒	☐	☒	☐	☒	☐	☒	☐	Logit	9, 12, 28
Cascetta et al. (2002)	☒	☐	☐	☒	☒	☒	☐	☐	☐	Binomial	1, 12, 17
Parkany et al. (2006b)	☐	☐	☒	☒	☐	☒	☒	☒	☐	Binary logit	3, 9, 10
Cools et al. (2009)	☒	☐	☐	☒	☐	☒	☐	☐	☐	Regression	1, 9, 10, 23
Papinski et al. (2009)	☐	☐	☒	☒	☐	☒	☒	☐	☐	Statistical analysis	1, 3, 4, 6, 21, 22
Ben-Elia and Shiftan (2010)	☐	☒	☐	☒	☐	☒	☐	☐	☐	Mixed logit	14, 15, 25, 29
Tawfik et al. (2010b)	☒	☐	☐	☒	☐	☒	☐	☒	☒	Statistical analysis	1, 6, 9, 15, 24, 25
Schlaich (2010)	☐	☐	☒	☒	☐	☐	☐	☐	☐	Logit model	3, 28
Prato et al. (2012)	☒	☐	☐	☒	☒	☒	☐	☒	☐	Latent variable	1, 5, 6, 7, 9, 13, 27
Kaplan and Prato (2012)	☒	☐	☐	☒	☐	☒	☐	☒	☐	Path-Size logit	1, 5, 6, 7, 9, 13, 27
Gan and Chen (2013)	☐	☒	☐	☒	☐	☒	☐	☐	☐	Binary logit	14, 15, 16
Jou and Yeh (2013)	☐	☒	☐	☒	☐	☒	☐	☐	☐	Mixed logit	1,5,6,8,10,12,18, 19
Tawfik and Rakha (2013)	☒	☐	☐	☒	☐	☐	☐	☒	☒	Latent class	1, 6, 9, 15, 24, 26, 27
Ramaekers et al. (2013)	☐	☐	☒	☒	☐	☒	☒	☐	☐	Multinomial logit	8, 9, 10, 11
Koller-Matschke et al. (2013)	☒	☒	☒	☒	☐	☐	☐	☐	☐	Statistical analysis	1, 18, 28, 29
Habib, Morency, Trépanier, and Salem (2013)	☒	☐	☐	☒	☐	☒	☐	☐	☐	Independent availability logit	1, 8, 10, 20
Vacca and Meloni (2014)	☐	☐	☒	☒	☐	☒	☐	☐	☐	Mixed logit	1, 5, 9, 12, 15, 22, 24
Hess, Quddus, Rieser-Schüssler, and Daly (2015)	☐	☐	☒	☒	☐	☐	☐	☐	☐	Path-Size logit	1, 6, 11, 19
Manley, Addison, and Cheng (2015b)	☐	☐	☒	☒	☐	☐	☐	☐	☐	Statistical analysis	20
Lai and Bierlaire (2015)	☐	☐	☒	☒	☐	☐	☐	☐	☐	Cross-Nested logit	5, 6, 22
Dalumpines and Scott (2017b)	☐	☐	☒	☒	☐	☒	☐	☐	☐	Path-Size logit	1, 5, 6, 7, 11
Alizadeh, Farooq, Morency, and Saunier (2017b)	☐	☐	☒	☒	☐	☐	☐	☐	☐	Nested and logit kernel	4, 5, 6, 20

***Factors affecting route choices**

1- Travel time	7- Number of turns	13- Delay	19- Toll rate / Cost	25- Learning process
2- Travel time reliability	8- Time of day	14- Network familiarity	20- Anchor points	26- Drivers' categories
3- Traffic conditions (level of service)	9- Socio-demographic	15- Driving experience	21- Stop signs	27- Personality traits
4- Number of segments	10- Trip purpose	16- Education	22- Traffic lights	28- Availability of information
5- Percentage of highway	11- Road type	17- Topology	23- Holidays	29- Type of information
6- Travel distance	12- Socio-economic	18- Choice inertia	24- Travel speed	30- Quality of information

[a] Observed Choices [b] Considered Choice Set [c] Demographics [d] Revealed Factors [e] Behavioural traits [f] Perception

CHAPTER 3 METHODOLOGICAL FRAMEWORK

The main objective of this thesis is to improve the understanding of drivers' route choices and revolves around several areas of improvement, namely modelling considerations, data collection framework, and the consideration set of route alternatives. A brief overview of the different components of route choice modelling has been presented in Chapter 2.

A combination of analytical and experimental research was needed to deliver the research objectives. This section summarizes some of the research gaps and existing limitations of the discussed route choice modelling approaches, which justify the significance of this research and its objectives. In the following subsections, the execution of each major part of the research plan along with the scope of the conducted research activities are elaborated.

3.1 Modelling Considerations

3.1.1 Hierarchy of Space and the Role of Anchor Points

As presented concisely in section 2.4, most of the proposed route choice models imply that individuals choose their routes based on route-level attributes, i.e. attributes concerning the whole trajectory, such as the total travel time, the total travel distance, the total number of turns, and the average travel speed.

However, in addition to the important role of route-level attributes, several studies have argued that anchor points influence route choice decisions (Couclelis, Golledge, Gale, & Tobler, 1987; Golledge, Smith, Pellegrino, Doherty, & Marshall, 1985; Habib et al., 2013; Kaplan & Prato, 2012; Lynch, 1960; Prato & Bekhor, 2007a; Prato et al., 2012). In route choice studies, anchor points are defined as prominent features along a route, such as major bridges, highways, interchanges, and intersections, with applications in cognitive tasks, such as way-finding, distance assessment, and direction estimation. It has been suggested that individuals have a hierarchical planning strategy following the hierarchical representation of space and its connectivity (Manley, Orr, & Cheng, 2015; Wiener & Mallot, 2003), which yields an anchor-based navigation in which individuals orient themselves based on distinguished features of the route (Foo, Warren, Duchon, & Tarr, 2005). However, most estimated route-based models do not take into account the effect of anchor-points and give an exclusive importance to route-level attributes, and existing link-based models allocate the same level of importance to every link regardless of its possible importance in drivers' route choice decisions. This limitation is addressed in the first article, entitled "On the Role of Bridges as Anchor Points in Route Choice Modelling" presented in Chapter 4 of this thesis.

In summary, this study explored the application of a nested structure to improve the behavioural aspect of route choice modelling by incorporating the effect of anchor points and space hierarchy on drivers' decision-making process. In the proposed anchor-based nested structure, the effect of route level attributes is incorporated in route choice decisions along with the effect of anchor points. First, we adopt a classic Nested Logit structure within a utility maximization discrete choice framework, in which upper nests correspond to anchor points and lower nests include route alternatives. Second, a nested Logit Kernel model is estimated to capture the reciprocal effect of route level attributes and anchor points on route selection. In the former model, the nested structure captures the shared unobserved components of the utility function among routes crossing the same bridge, while in the latter, the adopted factor analytic approach accounts for the interdependencies and latent similarities.

To evaluate the proposed modelling approaches, taxi trips between the islands of Montreal and Laval, Canada have been studied. These two cities are separated by a river and are connected through several bridges. These bridges are usually prone to become traffic bottlenecks and face recurrent congestion. Despite their small share in the whole route, they have a significant impact on the total travel time and are influential points in the process of route selection. Moreover, routes crossing the same anchor points share unobserved components such as safety, scenery, and driving comfort, emerging from the similarities of the road network and geographical characteristics. In our application, bridges are considered as anchor points for trips between Montreal and Laval, and their effects on route choice decisions have been evaluated in conjunction with route level attributes.

In general, this study illustrates that the proposed nested structures provide better estimates and validation performance compared to other route-based models, which clearly underscores the importance of considering the effects of anchor points in conjunction with route-level attributes. Moreover, the proposed models are easily manageable and practical, even by considering many alternatives or multiple anchor points.

3.1.2 Behavioural Classification

As discussed in Chapter 2, the RUM framework assumes that drivers tend to choose their routes in a way that maximizes their perceived utilities. Several factors might affect the perceived cost function, such as different levels of information, different capacities to process them, different computational and prediction abilities, network characteristics, travel possibilities, household characteristics, age, gender, and attitudes (Ben-Akiva et al., 1991; Bovy & Stern, 1990b). Moreover, satisfactory or unsatisfactory choice experiences can affect drivers' personal preferences towards a particular decision. It is argued that the repetition of satisfactory results can become a stable preference and drivers tend to repeat the same choices, forming various commute patterns (Ben-Akiva et al., 1991; Dia, 2002).

Understanding the behavioural heterogeneity in drivers' route choices is very important for city and transportation planners, and the stratification of a studied population according to their preferences would improve the effectiveness of transport measures and the efficiency of policy implications, since it better captures the heterogeneity across different segments of the population.

Chapter 5 empirically examines whether this behavioural heterogeneity can be observed in real world route choices. It also investigates the possibility of identifying different types of route choice behaviours among drivers. Behavioural classifications have been conducted on various segments of the population, such as pedestrians (Okamoto et al., 2011), bike users (Damant-Sirois, Grimsrud, & El-Geneidy, 2014; Dill & McNeil, 2013; Geller, 2009; Kroesen & Handy, 2014), bike-sharing members (Reinoso & Farooq, 2015), and carsharing members (Morency, Trepanier, & Agard, 2011). Although previous studies have shown that different categories of road users are observable, there is a lack of a representative classification of drivers' route choice behaviours, based on their actual choices over a long duration of time.

A total number of 1,746 taxi drivers' traces, comprising more than 22,000 trips, originating in Montreal with a destination in Laval, have been studied. Studies focusing on driving patterns over an extended period of time appear to be few, which might be mostly due to the challenges of data collection and analysis. The availability of large and rich datasets from taxi companies that maintain detailed GPS trajectories of their fleets for a long duration of time, opens up the possibility of such studies. Moreover, understanding taxi drivers' route choice behaviours helps to better comprehend urban traffic dynamics and is very important to the city and transportation planners

(Manley, Addison, & Cheng, 2015a). A Principal Component Analysis followed by a Hierarchical Agglomerative Clustering method was used to extract behavioural clusters. Four major types of route choice behaviour were observed, which showed significant variations based on the time of day and the travelled distance. The incorporation of different driver categories can improve the estimation and prediction accuracy of route choice models, and can be used in defining different functional forms for traffic assignment models (Parkany, Du, Aultman-Hall, & Gallagher, 2006a; Ramaekers et al., 2013; Tawfik, Rakha, & Miller, 2010c).

3.1.3 Behavioural Traits and Latent Heterogeneity

Following the discussion around factors affecting drivers route choices, these decisions might not be exclusively dependent on observable variables (such as traffic conditions, speed limits, number of turns, pavement quality, trip purpose, travel time, weather conditions, time of day, and traffic information, etc.), but also on latent variables, which cannot be directly observed, and measured, such as attitudes, perceptions, and lifestyle preferences and are considered to be intrinsically subjective (Gärling, Gillholm, & Gärling, 1998; Hurtubia, Nguyen, Glerum, & Bierlaire, 2014; McFadden, 1986, 1999; Raveau, Álvarez-Daziano, Yáñez, Bolduc, & de Dios Ortúzar, 2010). Moreover, different segments of the population, characterized by some of these latent constructs, might also have different choice behaviours (Hurtubia et al., 2014).

The explicit incorporation of these amorphous constructs has been mostly neglected in route choice models, and the latent behavioural heterogeneity among the population has mostly been ignored by assuming that all the individuals in the sample population have similar levels of driving experience, spatial knowledge, familiarity with the road network, ability to process information, motivation to compare all the considered alternatives, etc. Ignoring these sources of heterogeneity could reduce the explanatory power of the model and introduce forecasting errors (Ben-Akiva, Bolduc, & Bradley, 1993; Prato et al., 2012; Walker, 2001).

In Chapter 7, we present and discuss a comprehensive framework to explicitly incorporate latent behavioural constructs as well as a probabilistic segmentation of the population based on drivers' perceptions and preferences. In the presented article, we apply the proposed framework to compare the route choice behaviour of frequent versus occasional drivers. For this purpose, we evaluate the role of the underlying behavioural constructs on drivers' route choice decisions using the Integrated Choice and Latent Variable framework described by Walker (2001) and adapted to route choice studies by Prato et al. (2012). To properly incorporate the effect of segment heterogeneity and to distinguish between choice behaviours of the different classes of our sample population, we estimate the ICLV model within a Latent Class framework using a full information estimation approach (Bierlaire, 2016).

Data has been collected through a revealed preference web-based survey, conducted in Montreal, Canada in 2017. The survey has been designed to identify behavioural and attitudinal factors affecting drivers' route choice decisions. The modelling dataset for this study included 225 drivers residing and driving in the Greater Montreal Area (GMA). A thorough description of the data collection effort, recruitment methods, participants' characteristics, response rates, and dropouts is presented in Chapter 6. Results of this study confirmed that the inclusion of latent variables and latent heterogeneity across population segments significantly improves the explanatory power of the choice model.

3.2 Specialized Data Collection

As discussed, unobservable variables play a major role in route choice decisions, and various approaches have been developed to incorporate them into the modelling framework. Despite the attraction of such modelling frameworks, their application in route choice studies remains rare and occasional. This is mostly due to the fact that collecting behavioural data is cumbersome and time consuming (Sarkar & Mallikarjuna, 2017). Studies using revealed preference data are mostly based on either travel surveys or GPS data, where the presence of behavioural and attitudinal data is scarce. Moreover, studies based on stated preferences data mostly focus on observable attributes and avoid attitudinal questions to minimize respondents' burden.

In a two-stage route choice modelling process, defining a proper consideration set is a challenge. Usually, the set of route alternatives considered by the driver is not observed and available to the analyst. Therefore, the choice set generation algorithms discussed in Chapter 2 are adopted to create simulated consideration sets for each observation. It has been well established in the literature that the generated set should include alternatives that are attractive to the driver in a real world choice situation, and the misspecification of the size and composition of the considered choice set greatly affect model's estimates and may lead to erroneous predicted demand level (Bliemer & Bovy, 2008; Geda, 2014; Hess & Daly, 2010; Hoogendoorn-Lanser et al., 2005; Peters et al., 1995; Prato

& Bekhor, 2006, 2007b; Prato et al., 2012; Schuessler & Axhausen, 2009; Swait & Ben-Akiva, 1987a).

Therefore, in order to improve the estimation and prediction abilities of route choice models, it might be beneficial to first observe drivers' revealed preferences in real route choice situations, second, to identify behavioural and attitudinal factors as additional sources of heterogeneity affecting their decisions, and finally, to get a better grasp of the formation process, size and composition of drivers' considered sets of route alternatives.

Chapter 6 presents the development and implementation of the proposed revealed preference web-based survey, designed to observe drivers' revealed route choices towards their most frequently visited destinations, and to identify behavioural and attitudinal factors affecting them. Also, the survey aimed at observing drivers' consideration sets of route alternatives and characterizing them based on drivers' perceptions. Drivers residing and driving in the Greater Montreal Area (GMA), covering approximately 9840 square kilometers with a population of roughly 4 million inhabitants have been targeted (Transport, 2013). The survey has been prepared in both the French and English languages. In order to decrease the respondent burden, mitigate the implementation cost, and enhance the data quality, a high performance front-end user interface with an elaborated graphic design was adopted.

First, we collected typical information on sociodemographic and socioeconomic characteristics of participants. Then, respondents were asked to specify the destination point to which they drove most frequently and all the alternative routes which they considered for the specified trip. They were also asked to specify the frequency of the trip on a weekly basis. Additionally, they were requested to provide their level of agreement, on a five-point Likert scale, to a list of statements designed to reveal drivers' attitudes, preferences and perceptions towards choosing a route.

In total, 74 questions were asked and considering a 95th percentile threshold, the average response time was found to be around 16 minutes. By the end of the three-month data collection period, 843 individuals started the survey from which 539 completed it, while the remaining 304 dropped out. The data collected in this survey has been used in studies presented in Chapters 7 and 8.

3.3 Consideration Set of Route Alternatives

As previously discussed, considered sets of route alternatives are mostly unknown and unobserved. Their formation depends on objective constraints, such as route attributes (e.g. maximum number of turns, number of traffic signals, etc.), as well as subjective criteria, such as individuals' attitudes, perceptions and experiences (Ben-Akiva & Boccara, 1995). The way individual drivers derive their actual consideration sets of route alternatives, and factors affecting the size and composition of these choice sets is a complex and ongoing research issue (Hoogendoorn-Lanser et al., 2005; Prato et al., 2012; Schuessler & Axhausen, 2009).

In the last paper, presented in Chapter 8, we observed drivers' actual consideration sets of route alternatives and analysed the effect of different factors affecting their sizes. We used the data collected by the survey presented in Chapter 6, designed to collect detailed information on drivers' revealed consideration set of route alternatives. The data contains information regarding drivers' sociodemographic and socioeconomic characteristics, route specifications, as well as drivers' preferences and attitudes. The studied dataset is composed of 506 respondents residing and travelling in the GMA. A total of 988 route alternatives have been declared.

Studied factors have been classified into six broad categories, namely personal attributes, declared factors, behavioural indicators, incentives, awareness determinants, and spatial, temporal and environmental components. Four different clusters were defined based on the number of considered alternatives and the effect of each factors on each cluster was investigated. These four clusters were labelled 1) Determined Cautious Drivers, 2) Biased Habitual Drivers, 3) Middling Impartial Drivers, and 4) Swayable Conscious Drivers. The observed choice set may improve model's estimation and prediction efficiency by providing detailed information on travellers' preferences.

The next five chapters present the five articles discussing the above-mentioned aspects in more details. These articles are published or submitted for publication in scientific journals. Each one has an independent structure generally consisting of an abstract, introduction, background, methodology and data description, results and discussion, conclusions and references. A general discussion and comprehensive conclusions are then presented in Chapters 9 and 10, respectively.

CHAPTER 4 ARTICLE 1: ON THE ROLE OF BRIDGES AS ANCHOR POINTS IN ROUTE CHOICE MODELLING

Hamzeh Alizadeh[1*], Bilal Farooq[1], Catherine Morency[1], Nicolas Saunier[1]

Paper published in the *Transportation Journal*

First Online: 2 February 2017, pp: 1-26

DOI: https://doi.org/10.1007/s11116-017-9761-7

[1] Department of Civil, Geological and Mining Engineering, Polytechnique Montréal, C.P. 6079, succ. Centre-Ville Montréal, Canada

[*] Corresponding author

Abstract

This work builds upon the thought that individuals allocate higher levels of importance to some particular features of the route, so called anchor points. In this work we argue that the consideration of both route-level attributes and anchor points would enhance the behavioural aspect of route choice models as well as their estimation and prediction abilities. Global Positioning System (GPS) traces have been used to investigate the effect of bridges as anchor points for trips between Montreal and its Northern suburb, Laval. A classic Nested Logit and a nested Logit Kernel model have been estimated, in which interdependencies among routes crossing the same bridge are captured through the nested structure and the adopted factor analytic approach, respectively. A Metropolis-Hastings path-sampling algorithm is applied, for the first time, on a large road network with more than 40,000 nodes and 19,000 links to provide the consideration choice set. Estimates are then compared to three alternate models, representing route-based and anchor-based formulations; namely Path-Size Logit, Extended Path-Size Logit, and Independent Availability Logit models. Empirical results showed that the proposed nested structures with MH sampling provide better estimates and also perform better in the validation step with respect to comparative models. Findings underscore the importance of considering anchor points in conjunction with route level attributes in route choice decisions.

Keywords: Route choice, Bridge choice, Anchor points, Discrete choice models, Nested Logit, GPS, Metropolis-Hastings

4.1 Introduction

Route choice modelling is probably one of the most complex and challenging problems in traffic assignment. It investigates the process of route selection by an individual, making a trip between predefined origin and destination (OD) pairs. The heterogeneity in travellers' behavioural characteristics, in conjunction with the complex effect of route attributes, further increases the inherent complexity of route choice modelling. Although several approaches have been proposed to tackle this problem, one of the remaining challenges in route choice modelling is the consistency of the modelling approach with the underlying behavioural process of drivers' decision making.

In general, most of the proposed route choice models focus on route related attributes of choice alternatives. This implies that in these *"route-based"* formulations, the route is perceived as an entity, and only attributes concerning the whole trajectory are used to characterize each choice alternative. From a behavioural perspective, this formulation suggests that the consideration set is formed based on route-level characteristics of trajectories and the final choice is made by selecting an entire route out of a considered set of alternatives. C-Logit (Cascetta et al., 1996b) and Path-Size Logit (PSL) (Ben-Akiva & Bierlaire, 1999a) are among the most widely used route-based models. In these models, the similarity issue between alternatives has been addressed by adding a correction term to the deterministic part of the utility function, which alters the utility of paths[1] based on their similarities. However, the applied correction factors in these models account only for similarities between the considered set of paths. To overcome this limitation, Frejinger et al. (2009) have proposed the Extended-Path-Size Logit (EPSL) model, which accounts for correlations between sampled and non-sampled alternatives.

An alternate approach offers a *"link-based"* formulation to deal with the correlation issue among alternatives. This approach, firstly proposed by Vovsha and Bekhor (1998) and later applied by Lai and Bierlaire (2015), adopts a cross-nested logit structure in which each link of the network constitutes a nest and each route belongs to several nests. Another application of the sequential link choice method has been applied by Fosgerau et al. (2013) in the context of a recursive logit model, which can be consistently estimated without the need of path sampling. Since a real network

[1] The terms route and path are used interchangeably in this article.

consists of a large number of links, the estimation of these models could be very computationally expensive.

A third approach is the *"anchor-based"* formulation which gives an exclusive importance to the effect of some prominent features of the road network, so-called landmarks or anchor points. This approach basically argues that anchor points are decisive points based on which drivers choose their paths. It has been pointed out by several researchers that individual's perception of the road network follows a hierarchical representation (Hirtle & Jonides, 1985; Holding, 1994), and anchor points play a crucial role in the behavioural nature of route choice decisions (Habib et al., 2013; Kazagli & Bierlaire, 2015; Manley, Orr, et al., 2015). The space hierarchy and the role of anchor points have also been found to be significant in similar decision making contexts such as location choice modelling (Elgar, Farooq, & Miller, 2009; Elgar, Farooq, & Miller, 2015).

The importance of anchor points is emphasized in riverside cities. In these cities, the two sides, separated by the river, are usually connected through several bridges and tunnels. These infrastructures are usually prone to become traffic bottlenecks and to face recurrent traffic congestion (Habib et al., 2013; Sun, Wu, Ma, & Long, 2014; Woo et al., 2015). The high travel time variability of these small segments of the route, which is usually due to the large fluctuation of travel demand, turns them into influential points in the process of route selection. Accordingly, the choice of anchor points has a major effect on route selection, and routes crossing same anchor points share unobserved components such as safety, scenery, driving comfort, etc., emerging from the similarities of the road network and geographical characteristics.

In this work we argue that the three abovementioned formulations, in isolation, may not be behaviourally accurate enough to represent the underlying nature of route choice behaviour; rather a hybrid approach is needed. Route-based formulations suggest that route selection is mostly based on route-level attributes (Manley, Addison, et al., 2015a), and ignore the influence of anchor points. Link-based formulations neglect the higher importance of anchor points by allocating the same level of importance to every link. Moreover, the anchor-based model proposed by Habib et al. (2013) incorporates a full probabilistic choice set generation, which is behaviourally inaccurate, and theoretically impractical and unmanageable in large route choice datasets and real world networks (Prato, 2009b).

We propose a generic "anchor-based nested" structure to promote the behavioural aspect of route choice models by incorporating the effect of route level attributes as well as anchor points on drivers' choices. First, we adopt a classic Nested Logit (NL) structure within a discrete choice framework, in which upper nests correspond to anchor points and lower nests include route alternatives. Second, a nested Logit Kernel (LK) model is estimated to capture the reciprocal effect of route level attributes and anchor points on route selection. In the former model, the nested structure captures the shared unobserved components of the utility function among routes crossing the same bridge, while in the latter, the adopted factor analytic approach accounts for the interdependencies and latent similarities.

Similarly to anchor-based models, these approaches allocate a distinctive importance to the selection of anchor points as crucial segments of the route. Moreover, they can handle very large datasets and real world networks; considering multiple route alternatives within each bridge is easily manageable; and route-level attributes are also considered to be decisive and influential in the final route selection.

To explore the performance of the proposed formulations, GPS traces of taxi trips between the islands of Montreal and Laval have been used. The unique aspect of these trips is that drivers have to choose among a maximum of nine bridges separating the two regions. The access to these bridges face recurrent congestion, and despite their small share in the whole route, they have a significant impact on the total travel time and hence on drivers' route choice decisions. In our application, bridges are considered as anchor points for trips between Montreal and Laval, and their effects on route choice decisions have been evaluated in conjunction with route level attributes. A very large real-world road network, with more than 40,000 nodes and 19000 links, is used for choice set generation as well as model estimation. Estimates are then compared to three alternate models, representing route-based and anchor-based formulations; namely PSL, EPSL, and Independent Availability Logit (IAL) models. Taxi drivers are considered to have a more precise knowledge of the road network and its traffic conditions, due to their higher driving experience. In order to capture the effect of anchor-points, we have focused on the behaviour of taxi drivers as well-informed individuals who are more familiar with travel time variations and congestion periods over bridges connecting Montreal to Laval.

This work contributes to the existing state-of-the-art through the following aspects: (1) the presented formulations capture the effect of anchor points in conjunction with route level attributes in route choice modelling, (2) they improve the behavioural aspect of anchor-based formulations by capturing shared unobserved components among route alternatives crossing the same anchor point, and (3) the MH algorithm has been employed, for the first time, on a large real world route network to generate alternative choice sets.

This paper is organized as follows. First we review earlier approaches to route choice modelling and in that context further clarify the contributions of this study. The case study and data are presented next. We then discuss in detail the proposed econometric formulations and the estimated comparative models, as well as their respective utility function specifications and choice set generation algorithms. We then discuss the results, validation process, and comparison between models. In the end, we highlight the most significant findings of this study, underscore its limitations, and suggest further research directions.

4.2 State-of-the-Art

There is a large body of literature in microeconomics, behavioural science, psychology, and behavioural geography that focuses on improving the understanding of the underlying process of decision making. Accordingly, several modelling frameworks have been proposed to simulate drivers' route choice behaviour. Prospect theory (Gao et al., 2010; Kahneman & Tversky, 1979) and cumulative prospect theory (Connors & Sumalee, 2009; Tversky & Kahneman, 1992; Xu et al., 2011) have been applied by researchers to take into account the limited rationality of drivers in making decisions, by incorporating psychological and behavioural aspects. In some other studies, the uncertainty and imprecision of drivers in making route choice decisions have been taken into account using Fuzzy Logic (Henn, 2003; Luisa De Maio & Vitetta, 2015; Murat & Uludag, 2008; Quattrone & Vitetta, 2011). Artificial neural networks have also been used to take into account the non-linearity of the decision making process by imitating the human conscious structure (Dougherty, 1995; Kim et al., 2005). Recently, a Random Regret Minimization (RRM) approach has been adopted by Prato (2014) to model route choices. This approach argues that choice makers tend to choose the alternative that minimizes the regret of not having chosen other alternatives.

Among the proposed approaches, Random Utility Maximization (RUM) models have received considerable attention. In this approach, individuals' preferences are represented by a value called *"utility"*, which captures the effect of different factors on the actual choice, and decision makers tend to maximize their perceived utilities. Since the decision maker may not have a perfect knowledge about these factors, an error component has been introduced to take into account the stochasticity and imprecision caused by uncertainty and behavioural randomness (Ben-Akiva & Bierlaire, 2003). This paper is built upon the random utility maximization framework.

In the context of route choice, the perceived utility can be attributed to factors such as travel time, distance, congestion, safety, scenery, route complexity, number of traffic signals, trip purpose, fuel consumption, toll, road type, anchor points, etc. Behavioural and mental factors such as the inertia of taking the same route, memory, spatial abilities, driving experience, and the learning process may also play a role and add to the complexity of the modelling process (Prato, 2009b).

It has been suggested that individuals have a hierarchical planning strategy following the hierarchical representation of space and its connectivity (Manley, Orr, et al., 2015; Wiener & Mallot, 2003), which yields an anchor-based navigation in which individuals orient themselves based on distinguished features of the route (Foo et al., 2005). Anchor points are defined as being important focal points and cognitively salient cues with prominent features, with applications in cognitive tasks, comprising way-finding, distance assessment, and direction estimation. Major route infrastructures, such as bridges, highways, eminent road interchanges, intersections and roundabouts, can be considered as anchor points in route choice modelling.

Several studies have argued that anchor points influence route choice decisions (Couclelis et al., 1987; Golledge et al., 1985; Habib et al., 2013; Kaplan & Prato, 2012; Lynch, 1960; Prato & Bekhor, 2007a; Prato et al., 2012). Manley, Addison, et al. (2015a) studied minicab drivers in London, using GIS and statistical analysis, and concluded that their route choice behaviour is poorly described by shortest-path algorithms and is improved when the role of anchor points is considered. To emphasize the significant role of anchor points, they have also proposed a conceptual subjective anchor-based route choice modelling schema. Similarly, Kazagli and Bierlaire (2015) argue that drivers describe their routes using a short sequence of Mental Representation Items (MRIs) such as anchor points or pieces of infrastructures instead of using a link-sequence representation. Moreover, a recent study by Manley, Orr, et al. (2015) confirms that the mental representation of the spatial hierarchy influences route choices. They propose a coarse to granular hierarchical representation of space, represented from top to bottom by *Regions*, *Nodes*, and *Roads*, where *Regions* represent clusters of nodes sharing a common characteristic; *Nodes* represent certain road junctions, landmarks, and anchor points; and *Roads* form the basis of the hierarchy, defining the route between consecutive *Nodes*. In this hierarchical schema, route choice is made through the selection of a sequence of regions, nodes across subsequent regions, and eventually, roads between successive nodes.

Despite the undeniable importance of anchor points on drivers' route choice decisions, relatively little attention has been given to *anchor-based* route choice models. An interesting approach has been investigated by Habib et al. (2013) in which the authors applied an Independent Availability Logit (IAL) model, originally proposed by Swait and Ben-Akiva (1987b), in a route choice context to emphasize the role of bridge choice in route choice decisions; the case study was the Greater Montreal Area. The IAL model follows the probabilistic two-stage choice model proposed by Manski (1977) in which the selection probability of an alternative depends on the selection probability of all subsets of the universal choice set containing that particular alternative. The IAL model jointly estimates the final choice and choice sets among all the possible combinations.

The study by Habib et al. (2013) was based on data collected from the OD survey of Montreal, in which the authors had only access to declared chosen bridges. They have considered a shortest path algorithm, based on segments' speed limits, to generate one path per bridge for each OD pair, comprising the choice set of route alternatives. A noticeable limitation, pointed out by Prato (2009b), of the fully probabilistic choice set generation approach adopted by the IAL approach, is its immense calculation burden when the size of the choice set increases. Considering *m* as the total number of possible alternatives, the total number of possible non-empty choice sets $(2^m - 1)$ increases exponentially with *m*, which makes it impractical to apply this model in route choice problems, where the considered choice set is large. In Habib et al. (2013), the inability of the IAL model in handling large choice sets might be an additional reason to their data limitation issue for considering only one alternative per bridge. In behavioural terms, the consideration of a single shortest path per bridge scales down the route choice problem to a bridge choice problem. Moreover, it is behaviourally unrealistic to assume that decision makers would consider every possible subset of the consideration choice set, before making a choice. Also, it is worth mentioning that the problem of shared segments between alternatives is not addressed in this recent application of the IAL model.

4.3 Context and Dataset

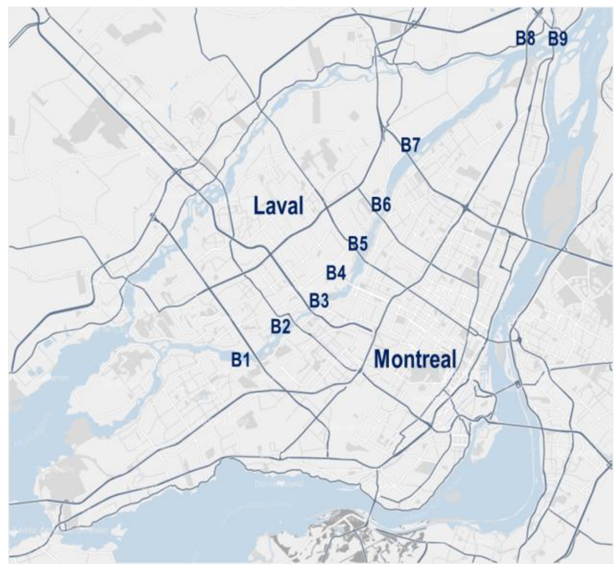
Nowadays, the prevalent use of GPS technology provides researchers with an abundance of high-resolution geospatial data, which allows obtaining continuous and detailed (link-by-link) information on drivers' travel paths, accompanied by possible additional information such as travel direction and speed. A relatively new source of GPS data is recorded by taxi companies around the world mainly for operational purposes.

This study is based on GPS traces of taxi drivers, collected by a taxi company, in the context of the metropolitan region of Greater Montreal, depicted in Figure 4.1(a). The data was collected by a taxi company that constitutes around 25% of the Montreal Island taxi fleet, and its operation is restricted to trips starting or ending in the central part the island. Data has been stored in a PostgreSQL database, and the PostGIS spatial extension has been added to support geographical datatypes and queries. A direction-based nearest link point-to-curve map matching algorithm has been adopted to associate each GPS record to the road network. In point-to-curve map matching algorithms, every GPS point is matched onto the closest link in the network. The major shortcoming of these algorithms is that they do not produce reliable results in high road density networks and specially at intersections due to directional problems (White, Bernstein, & Kornhauser, 2000; Zhou & Golledge, 2006). To overcome this issue, we associated each GPS record to its nearest link on the network with respect to its azimuth, so that it ensures that GPS points are not incorrectly matched to closer links with incorrect directions. A distance-based shortest-path algorithm has then been applied between consecutive GPS records to deduce the entire path for each trip.

Montreal is an island city, separated from its suburbs by two rivers. This means that drivers entering or exiting Montreal need to cross one of the sixteen bridges connecting Montreal to its suburbs. These bridges face recurrent congestion and act as bottlenecks. For trips heading to or exiting Montreal, bridge choice can have a significant importance on route choice decisions (Habib et al., 2013). Considering bridges as anchor points, the geographical context of Montreal allows us to study the effect of anchor points in conjunction with route level attributes in route choice decisions.

In this study we focus on trips taking place between the Islands of Montreal and Laval, the largest suburb of Montreal located on the north of the city. These two islands are directly connected through seven bridges; B1 to B7. Since bridges B8 and B9 might also provide convenient alternatives for trips between the East of Montreal and Laval, they have been included in this study. Figure 4.1(b) depicts the location of these nine bridges and Table 4.1 summarizes some of the pertaining properties of these bridges. Montreal and Laval cover a total surface of 632.3 km^2 containing a population of roughly 2.3 million inhabitants (Communauté métropolitaine de montréal, 2012). Their road networks comprise more than 40,000 nodes and 19,000 links. The network data has been extracted from OpenStreetMap project in the format of geographical layers.

(a)



(b)

Figure 4.1: Context of the studied region. (a) The metropolitan region of Greater Montreal; (b) Locations of bridges connecting Montreal to Laval

Table 4.1: Description of bridges connecting Montreal to Laval

ID	Name	No. of lanes	Length* (km)	Road type	Speed limit	Toll
B1	Pont Louis-Bisson	4	0.535	Highway (13)	100	No
B2	Pont Lachapelle	3	0.264	Arterial	50	No
B3	Pont Mederic-Martin	4	0.361	Highway (15)	100	No
B4	Pont Viau	2	0.658	Arterial	50	No
B5	Pont Papineau-Leblanc	3	0.425	Highway (19)	100	No
B6	Pont Pie-IX	3	0.658	Arterial	80	No
B7	Pont Olivier-Charbonneau	3	1.200	Highway (25)	100	Yes
B8	Pont Charles-De-Gaulle	3	1.450	Highway (40)	100	No
B9	Pont Le Gardeur	2	0.500	Arterial	100	No

* Bridge lengths have been measured in QGIS software.

GPS records for the month of October 2014 have been extracted for this study. The dataset includes two tables: a) *"GPS"* table containing information for every recorded point such as ID, time, geographical position, speed, and etc.; b) the *"Events"* table, including information regarding the state of the taxi such as when the passenger has boarded and got off the taxi, etc. These two tables are related through a unique trip identifier ("Ride ID"). Table 4.2 shows the schema of the dataset.

Table 4.2: Internal structure of the database

GPS Table

- ID
- Time
- Object ID
- Longitude
- Latitude
- Status
- Speed
- Direction
- Employer ID
- **Ride ID**

Events Table

- **Ride ID**
- Events:
  - Flag
  - Accept
  - Reject
  - Cancel
  - Meter ON/OFF
  - Payment
  - …

Our dataset consists of 4409 GPS records comprising 543 journeys with an average length of around 11 km and a standard deviation of 9 km.The dataset comprises weekdays, as well as weekends and trips made in peak hours as well as off-peak trips. Eighty percent of the dataset (434 records) was randomly selected for calibration purposes, while the remainder twenty percent (109 records) was used for result validation.

4.4 Methodology

The principal aim of this research is to provide a behavioural framework, which explicitly takes into account the effect of anchor points as well as route-level attributes in route selection. To highlight the importance of anchor points in conjunction with route-level attributes, we adopt the following nested structures. First, we adopt a classic Nested Logit model in which the effect of anchor points is addressed in upper nests while route level decisions are represented in lower nests. Then, we adopt a nested Logit Kernel model, which accounts for the interdependencies of route alternatives crossing the same anchor point through the specification of its error structure.

In behavioural terms, our application of these nested structures suggests that individual taxi drivers, travelling between Montreal and Laval, consider bridges as crucial elements, along with other route level attributes, which affect their route choices. This section delineates the formulations of the abovementioned econometric models, presents the comparative route-based and anchor-based models, and describes their respective utility functions and choice set generation algorithms.

4.4.1 Econometric model formulation

4.4.1.1 Nested Logit

The Nested Logit (NL) formulation was proposed by Ben-Akiva (1973) and proved to be consistent with stochastic utility maximization theory by McFadden (1978). It is an extension of the Multinomial Logit (MNL), and captures some of the unobserved similarities among alternatives by dividing the choice set into several nests. These nests are considered to be collectively exhaustive and mutually exclusive in covering the considered alternatives. Every nest contains a subset of alternatives sharing a particular characteristic, independent from other subsets of alternatives in other nests. In other words, the probability of choosing an alternative from a nest is considered to be independent from alternatives in other nests, which is known as the Independence of Irrelevant Alternatives (IIA) property.Therefore, the probability of choosing an alternative can be expressed as the product of the conditional probability of choosing that alternative given a particular nest and the choice probability of that respective nest (Ben-Akiva, 1973; Guevara & Ben-Akiva, 2013). It is worth mentioning that in our case, the nine considered bridges between Montreal and Laval are located relatively far apart from each other, and it is not far-fetched to assume that the IIA property holds within each nest of the model.

NL is a member of the Generalized Extreme Value (GEV) models (McFadden, 1978), also known as Multivariate Extreme Value (MEV) models (Guevara & Ben-Akiva, 2013). Within this framework, the probability of choosing alternative i by individual n within the true choice set C_n is given by

$$P(i|C_n) = \frac{e^{V_{in}} G_i(e^{V_{1n}}, \dots, e^{V_{J_n}})}{G(e^{V_{1n}}, \dots, e^{V_{J_n}})} \qquad (4.1)$$

$$G_i(e^{V_{1n}}, \dots, e^{V_{J_n}}) = \frac{\partial G}{\partial e^{V_{in}}}(e^{V_{1n}}, \dots, e^{V_{J_n}}) \qquad (4.1)$$

where G is a non-negative differentiable MEV generating function, G_i is its partial derivative with respect to $e^{V_{in}}$, V_{in} specifies the systematic part of the utility function, and J_n is the number of alternatives in C_n. The probability of choosing alternative i from the true choice set is given by:

$$P(i|C_n) = \frac{e^{V_{in}+\ln G_i(e^{V_{1n}},\dots,e^{V_{J_n}})}}{\sum_{j=1}^{J_n} e^{V_{jn}+\ln G_j(e^{V_{1n}},\dots,e^{V_{J_n}})}} \qquad (4.2)$$

where the partial derivative of the MEV generating function for Nested Logit G_i, for the true choice set C, is:

$$G_i(C) = G_i(e^{V_1}, \dots, e^{V_J}) = \mu e^{V_i(\mu_m - 1)} \left(\sum_{i=1}^{J_m} e^{\mu_m V_i} \right)^{\frac{\mu}{\mu_m} - 1} \qquad (4.3)$$

in which μ and μ_m are scale parameters for the model and its nests, respectively, where $\mu/\mu_m \leq 1$, and m is the nest including alternative i. Since it is not feasible to enumerate the true choice set, a subset D has to be sampled which must include the chosen alternative i. To consistently estimate this model on a subset of alternatives, the correction approach proposed by McFadden (1978) can be adopted, in which an alternative specific correction term is added to the utility function.

$$P(i|D) = \frac{e^{V_{in}+ln\, G_i(C)+ln\, \pi(D|i)}}{\sum_{j=1}^{J} e^{V_{jn}+ln\, G_j(C)+ln\, \pi(D|j)}} \qquad (4.4)$$

$ln\, \pi(D|i)$ is the sampling correction factor, and $\pi(D|i)$ is the conditional probability of choosing subset D given the alternative i has been chosen. The approach developed by McFadden (1978) has been adopted by Bierlaire et al. (2008) to demonstrate that the maximization of the quasi-log-likelihood function of Eq. (4.4) yields consistent parameter estimates. It is worth mentioning that in our case, since a finite set of anchor points (bridges B1 to B9, connecting the two regions) is considered for the upper nest level, the application of this correction factor is found to be superfluous and unneeded.Although this function leads to the conditional probability of choosing alternative i given the subset D, it's application is not valid with sampling of alternatives because $\ln G_j(C)$ is still dependent on the true choice set (Guevara & Ben-Akiva, 2013). In order to compensate for the loss of information due to sampling in each nest, an expansion factor w should be considered to approximate the generating function based on the considered sample D'. Moreover, to take into account the physical overlap between routes crossing the same anchor point, the Extended Path-Size factor (see Eq. (4.16)) has been added to the deterministic part of the utility function, so that:

$$P(i|D,D',w) = \frac{e^{V_i + lnEPS_i + \ln\hat{G}_i(D',w) + \ln\pi(D|i)}}{\sum_{j=1}^{J} e^{V_j + lnEPS_j + \ln\hat{G}_j(D',w) + \ln\pi(D|j)}} \tag{4.5}$$

The closed form quasi-log-likelihood function has the following structure:

$$QL_{MEV,D,D',w} = \sum_{n=1}^{N} lnP(i|D,D',w) = \sum_{n=1}^{N} ln \frac{e^{V_i + lnEPS_i + \ln\hat{G}_i(D',w) + \ln\pi(D|i)}}{\sum_{j=1}^{J} e^{V_j + lnEPS_j + \ln\hat{G}_j(D',w) + \ln\pi(D|j)}} \tag{4.6}$$

Guevara and Ben-Akiva (2013) demonstrated that in order to achieve unbiasedness and consistency, the expansion factor should have the following structure:

$$w_j = \frac{k_j}{E[k_j]} \tag{4.7}$$

in which k_i is the number of times alternative i has been sampled, and $E[k_j]$ denotes its expected value or its sampling probability. In this study, we adopt the formulation proposed by Lai and Bierlaire (2015) to approximate Eq. (4.4) over the choice set D':

$$w_j = \frac{k_j}{k_s}\frac{b_{(s)}}{b_{(j)}} \tag{4.8}$$

where s denotes the path that has been sampled the most, k_s is the number of times alternative s has been sampled ($k_s \geq k_i \;\; \forall\, i \in D'$), and $b_{(s)}$ and $b_{(j)}$ are the theoretical frequencies of the most sampled path and path j, respectively.

4.4.1.2 Logit Kernel (LK)

Logit Kernel, which is a combination of Probit and Logit models, was first proposed by Bolduc and Ben-Akiva (1991) .The random component of its utility function is composed of a Probit-like term, which captures the interdependencies among alternatives, and an i.i.d. Gumbel distributed random component. The interdependencies between alternatives can be explicitly specified using a factor analytic approach, proposed by McFadden (1984). This approach accommodates different error structures and reduces the estimation complexity of the model (Bekhor et al., 2002; Bierlaire & Frejinger, 2005). The utility function for individual n is defined as below:

$$U_n = X_n\beta + F_n T\zeta_n + \nu_n \qquad (4.9)$$

where U_n - $(J_n \times 1)$ is the utility vector, and J_n is the number of alternative in the choice set C_n; X_n - $(J_n \times K)$ is the matrix of explanatory variables; β - $(K \times 1)$ is the vector of unknown parameters; F_n - $(J_n \times M)$ is the factor loading matrix; T - $(M \times M)$ is a diagonal matrix of the standard deviation of each factor; ζ_n - $(M \times 1)$ is the vector of i.i.d. random variables with zero mean and unit variance; and ν_n - $(J_n \times 1)$ is the vector of i.i.d. Gumbel distributed random term with zero location, a scale equal to μ, and a variance equal to $(\pi^2/6\mu^2)$.

The LK model can replicate any error structure and approximate any random utility model (Ben-Akiva et al., 2001; McFadden & Train, 2000; Walker et al., 2004). In a Nested Logit analog of the LK model, also known as the nested LK model, F_n is defined to be the alternative-nest incident matrix and is obtained by defining a dummy variable for each nest that equals 1 if an alternative belongs to that particular nest, and 0 otherwise. Moreover, ζ_n is usually assumed to be normally distributed $N(0,1)$, and T captures the amount of correlation between alternatives belonging to the same nest (Train, 2009; Walker et al., 2004). In this study, the correlation related to the physical overlap between routes crossing the same bridge is expressed through the Extended Path-Size factor (see Eq. (4.16)), which is added to the utility function ($lnEPS$). If the factors ζ_n are known, the probability of choice i given ζ_n is estimated using the MNL formulation:

$$\Lambda(i|\zeta_n) = \frac{e^{\mu(X_{in}\beta + lnEPS_{in} + F_{in}T\zeta_n)}}{\sum_{j=1}^{J_n} e^{\mu(X_{jn}\beta + lnEPS_{jn} + F_{jn}T\zeta_n)}} \qquad (4.10)$$

Since ζ_n is unknown, the unconditional probability takes the following form:

$$P(i) = \int_{\zeta} \Lambda(i|\zeta_n) \prod_{m=1}^{M} \phi(\zeta_m) d\zeta \qquad (4.11)$$

where $\phi(\zeta_m)$ is the standard univariate normal density function, and $\prod_{m=1}^{M} \phi(\zeta_m)$ represents the joint density function of ζ. Since the probability function does not have a closed form, it is approximated through simulation:

$$\hat{P}(i) = \frac{1}{D} \sum_{d=1}^{D} \Lambda(i|\zeta_n^d) \qquad (4.12)$$

where D is the number of simulation draws and ζ^d denotes draw d from the distribution of ζ. In this study, the factor analytic specification takes into account the effect of anchor points on route choice decisions and corresponds to bridges connecting Montreal to Laval. These factors capture the unobserved similarities among routes crossing the same anchor points. Accordingly, the F_n matrix is defined to be the route-bridge incident matrix with a dummy variable for each bridge, equal to 1 if a route crosses that particular bridge, and 0 otherwise.

4.4.2 Comparative models specification

The two presented anchor-based nested formulations are compared with three other models representing route-based and anchor-based formulations, namely the PSL, EPSL and IAL models. A concise introduction to these models follows.

4.4.2.1 Path-Size Logit

First, a PSL model is estimated as an instance of route-based models (Ben-Akiva & Bierlaire, 1999a; Prato, 2009b). This model uses a correction factor in the deterministic part of the utility function to account for the correlation among sampled paths; however, it ignores the correlation with non-sampled paths:

$$P_{PSL}(i|C_n) = \frac{e^{\mu(V_{in} + lnPS_{in})}}{\sum_{j \epsilon C_n} e^{\mu(V_{jn} + lnPS_{jn})}} \qquad (4.13)$$

where $P_{PSL}(i|C_n)$ is the conditional probability of user n choosing alternative i from the universal choice set C_n, μ is a scale factor, and V is the deterministic part of the utility function. The Path-Size factor $lnPS$ is added in a logarithmic scale to the deterministic part and is calculated as below:

$$PS_{in} = \sum_{a \in \Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j \epsilon \varphi_n} \delta_{aj}} \tag{4.14}$$

where L_a and L_i represent the length of link a and path i, Γ_i is the set of road segments in path i, φ_n denotes the considered choice set, and δ_{aj} is the link-path incident binary variable which is 1 if link a is on path i, and 0 otherwise. In other words, $\sum_{j \epsilon \varphi_n} \delta_{aj}$ indicates the total number of alternatives in the choice set sharing link a, for observation in φ_n.

4.4.2.2 Extended-Path-Size Logit

Second, an EPSL model (Frejinger et al., 2009) has been estimated which is also an instance of route-based models, in which the PS factor has been extended to take into account the correlation of each alternative with all the possible paths in the true choice set. However, the structure of the conditional probability stays the same:

$$P_{EPSL}(i|C_n) = \frac{e^{\mu(V_{in} + lnEPS_{in}) + ln(\frac{k_{in}}{q(i)})}}{\sum_{j \epsilon C_n} e^{\mu(V_{jn} + lnEPS_{jn}) + ln(\frac{k_{jn}}{q(j)})}} \tag{4.15}$$

and the EPS factor is defined by

$$EPS_{in} = \sum_{a \in \Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j \epsilon \varphi_n} \delta_{aj} \omega_{jn}} \tag{4.16}$$

where ω_{jn} is an extension factor with a value equal to 1 if $\delta_{aj} = 1$ or $q(j)R_n \geq 1$, and $1/(q(j)R_n)$ otherwise; where R_n denotes the total number of paths drawn with replacement from the universal choice set, $q(j)$ is the sampling probability of path j, and k_{jn} is the empirical frequency or the actual number of times path j is drawn.

4.4.2.3 Independent Availability Logit

Third, an IAL model (Swait & Ben-Akiva, 1987b) has been estimated to illustrate the performance of anchor-based models. In this formulation the choice set is latent and the probability of considering any combinations of alternative as the final choice set is calculated. The conditional probability of alternative i being chosen is calculated as

$$P_i^{IAL} = \sum_{D \subseteq \Gamma} P_D P_{i|D} = \sum_{D \subseteq \Gamma} \frac{\prod_{i \epsilon D} A_i \prod_{i \epsilon C-D} (1 - A_i)}{1 - \prod_{i \epsilon C} (1 - A_i)} \left(\frac{\exp(\mu V_i)}{\sum_{j \epsilon D} \exp(\mu V_j)} \right) \qquad (4.17)$$

where P_D is the probability of drawing the choice set D from a set of all possible non-empty choice sets Γ of the universal choice set C; $P_{i|D}$ denotes the probability of choosing alternative i from the choice set D; and $A_i = (1 + \exp(-\alpha x))^{-1}$, where x denotes attributes and α refers to parameters to be estimated. In order to achieve the proposed formulation for P_D, it is assumed that the IIA property holds for alternatives in the considered choice set.

4.4.3 Utility function specification

Four attributes are used to specify the systematic part of the utility function:

- Mtl_Len specifies the portion of trip length made on the island of Montreal,
- Lvl_Len denotes the portion of trip length made on the island of Laval,
- Hgw_Len stands for the portion of trip length made on highways, and
- Seg_Len indicates the average length of road segments[2].

The minimum, maximum, average and median values of these attributes over the whole dataset are reported in Table 4.3.

Table 4.3: Statistics on attributes included in the systematic part of the utility function

Attributes	**Min.**	**Max.**	**Avg.**	**Med.**
Mtl_Len (m)	266.6	22846.3	7459.9	6347.6
Lvl_Len (m)	117.5	23474.8	3729.0	2414.6
Hgw_Len (m)	0.0	33694.9	7381.6	3956.1
Seg_Len (m)	89.0	621.3	231.2	195.5

In estimating NL, LK, PSL, and EPSL models, the utility function for observation i is defined by:

---

[2] Road segments are defined to be the portion of a road between two consecutive junctions.

$$V_i = \beta_{Mtl\_Len} \times \text{Mtl\_Len}_i + \beta_{Lvl\_Len} \times \text{Lvl\_Len}_i + \beta_{Hgw\_Len} \times \text{Hgw\_Len}_i + \beta_{Seg\_Len} \times \text{Seg\_Len}_i \quad (4.18)$$

For the IAL model, the length of the trip made on the island of Montreal, which practically specifies the distance from the origin to the bridge, is used in the first part of the model to determine the selection probability of each choice sets:

$$A_i = \left(1 + \exp\left(-\beta_{Mtl\_Len} \times \text{Mtl\_Len}_i\right)\right)^{-1} \quad (4.19)$$

The other three variables are used to define the systematic part of the utility function:

$$V_i^{IAL} = \beta_{Lvl\_Len} \times \text{Lvl\_Len}_i + \beta_{Hgw\_Len} \times \text{Hgw\_Len}_i + \beta_{Seg\_Len} \times \text{Seg\_Len}_i \quad (4.20)$$

4.4.4 Choice set generation

The consideration set should include attractive alternatives. Since random sampling of alternatives in large universal choice sets is not efficient in terms of providing information, an *importance sampling* method would be more convenient and favorable (Hess & Daly, 2010). Several deterministic and probabilistic path generation methods, which are mostly based on repeated shortest path algorithms have been proposed in the literature to form the consideration set. Among them are link labelling (Ben-Akiva et al., 1984), link elimination (Azevedo et al., 1993), and link penalty (de la Barra et al., 1993). A thorough review of these methods is presented in (Prato, 2009b) and (Frejinger & Bierlaire, 2010). The major downside of using these methods is that they do not provide researchers with sampling probabilities of the generated alternatives. Model estimates based on these path generation methods are biased, unless the sampling probability of every alternative in the universal set is equal, which is not the case in route choice modelling.

Several alternative approaches have been proposed. Cascetta et al. (2002) adopted a two stage process, where in the first step, a complete set of alternatives is generated for all the observations by maximizing a coverage factor between the generated set and the set of routes perceived as available. Then, in the second step, a binomial Logit model is adopted to estimate the probability of including a given route in the users' consideration set. Frejinger et al. (2009) applied a biased random walk to sample a subset of paths and derived a sampling correction to obtain unbiased parameter estimates. More recently, Flötteröd and Bierlaire (2013) used a Metropolis-Hastings (MH) algorithm to generate sample sets based on an arbitrary distribution providing the sampling probability of each alternative. This algorithm requires a road network and a definition of path weight as an input. It uses an underlying Markov Chain process to sample alternatives and calculates its sampling probability without the need of normalizing it over the full choice set.

For the NL model estimated in this study, the MH algorithm has been adopted to generate nine alternatives per nest. In order to apply this algorithm on the large road network of Montreal and Laval, 100 separate input files have been prepared to provide the possibility of parallel calculation. Files have been imported into a cluster of 26 computers (2 processors Intel(R) Xeon(R) X5675 @ 3.07GHz) which took about 104 hours (4 days and 8 hours) to generate the output files. For the LK, PSL, and EPSL models, the MH algorithm has been adopted to draw 19 choice alternatives from the universal choice set. Similarly, to the NL model, 100 separate input files have been prepared to provide the possibility of parallel calculation, which resulted in a calculation time of 46 hours (1 day and 22 hours).

For the IAL model, a shortest path algorithm, using segments' speed limits as travel cost, has been adopted. The same algorithm was used in the original application by Habib et al. (2013) and provides the possibility of comparison between the outputs. In the IAL formulation, the feasible choice set built in the choice generation step is considered to be the equivalent of the universal choice set C. Since the algorithm calculates the choice probability of every non-empty subset of the universal choice, the number of considered alternatives should be restricted for computational purposes. The chosen alternative and eight shortest-distance paths, crossing the eight alternative bridges, comprise the nine feasible alternatives for each observation. Although Habib et al. (2013) showed that the IAL performs well in anchor choice prediction, it is very computationally expensive for applications involving large choice sets, such as route choice modelling applications.

In this study, different sizes of choice sets have been generated for practicality reasons. Considering larger choice sets would have increased the computational time dramatically; however, similarly to Elgar et al. (2009) the estimation gain would have probably been minor. In all the above mentioned applications, the chosen alternative has been added to the choice set, where it was not generated by the adopted choice set generation algorithm (Arifin, 2012; Dhakar & Srinivasan, 2014; Elgar et al., 2009; Frejinger et al., 2009; Habib et al., 2013; Hess et al., 2015; McFadden, 1978; Prato et al., 2012)

4.5 Results and discussion

This section presents, compares and discusses the estimation and prediction abilities of the aforementioned models. The BIOGEME software package (Bierlaire, 2003; Bierlaire & Fetiarison, 2009) has been used for all model estimations.

4.5.1 Estimation

Eighty percent of the observations, that is 434 trips, were randomly selected for estimation purposes. A heat-map of origin and destination points, presented in Figure 4.2, illustrates higher density regions, located mostly around metro stations, airport, downtown Montreal, and commercial centers, which have an expectedly higher taxi demand. A heat-map of chosen routes between OD pairs is illustrated in Figure 4.3 Around 40 % of the whole network size, considered for choice set generation, has been covered by drivers' route choices.

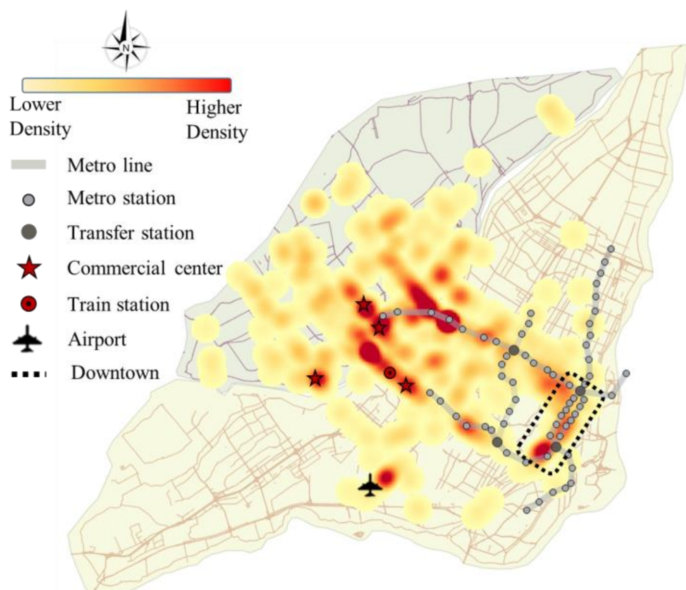


Figure 4.2: Heat-map of origin and destination points for trips between Montreal and Laval

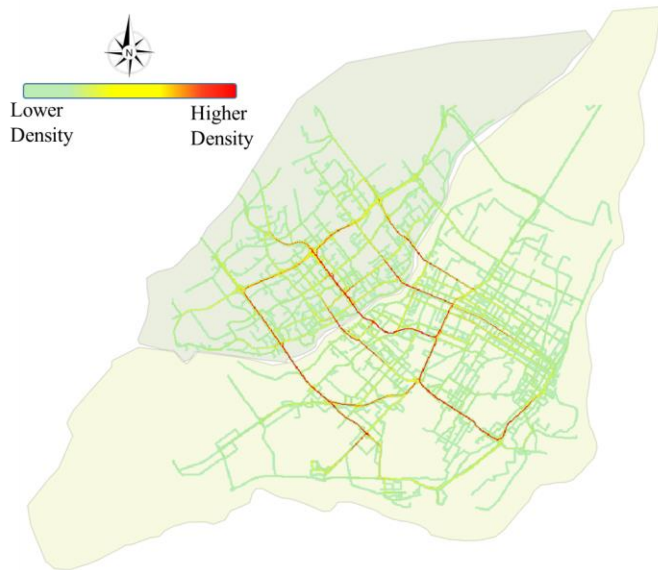


Figure 4.3 : Heat-map of chosen routes between OD pairs for trips between Montreal and Laval

A detailed description of models' estimates is provided in Table 4.4. Presented models are fully identified according to the smallest singular value approach implemented in BIOGEME (Bierlaire, 2015). The scale parameter in PSL, EPS and LK models were estimated while they have been fixed to 1 in IAL and NL models for identification purposes. Scale factors for nests in the NL model have been estimated and $\mu \leq \mu_m$ holds for every nest. Since the composition of the choice set differs from a model to another, not much can be inferred from the comparison of their scale factors. However, an out-of-sample validation method has been used to properly compare the models' performances, which will be presented and discussed in the subsequent section.

Intuitively, taxi drivers are apt to minimize their travel distance by choosing a shorter route, and their travel time by riding on segments with higher speed limits. This behaviour is confirmed by the obtained results from all the estimated models. Coefficients β_{Mtl_Len} and β_{Lvl_Len} are negative while β_{Hgw_Len} has a positive sign, meaning that taxi drivers are more willing to take shorter alternatives and are more inclined to ride on highways. These findings are in agreement with results reported by Duan and Wei (2014) who claimed that most taxi drivers tend to minimize their travel time. The effect of the average length of the segment is expectedly positive for PSL, EPSL, LK, and NL models, implying that taxi drivers tend to avoid intersections and prefer to take routes with a longer average segment length. However, this estimate has a negative sign for the IAL model, which might be attributed to the fact that alternatives are assumed to be completely independent from each other and their correlations have been neglected. The positive signs of β_{PS} and β_{EPS} are a negative correction of the utility for overlapping routes, giving a higher chance to less similar alternatives to be chosen. Similar findings are reported by (Bierlaire & Frejinger, 2008; Dhakar & Srinivasan, 2014; Prato & Bekhor, 2006, 2007a).

Note that the anchor-based nested models, namely LK and NL models, result in significantly higher Rho-square values compared to the route-based and anchor-based models. This emphasizes the importance of bringing the concept of anchor points in route choice modelling, so that it becomes more in line with the actual behaviour of drivers. The σ estimates are highly significant (except for bridge 6) for the LK model, implying that the factor analytic structure captures a significant correlation structure between routes crossing the same bridges. It is also noted that this effect is statistically different from the effect captured by the EPS factor. This is consistent with findings in Bekhor et al. (2002) and Bierlaire and Frejinger (2005), where the authors presented a LK route choice model considering subpath components. The better fit of the LK model with an EPS attribute over the PSL and EPSL models is in line with findings reported by Ramming (2002) and Bierlaire and Frejinger (2005). It is worth mentioning that the travel time-based shortest path algorithm, used in the choice set generation step of the IAL model, does not provide the researcher with the sampling probability of paths in order to correct the sampling effect. The better fit of all other models can be partially attributed to the application of MH algorithm, which provides the possibility of considering the sampling correction factor.

The estimation time has also been reported in Table 4.4. All model estimations were conducted on a machine with a core i7-4720HQ CPU running at 2.6GHz and a Random Access Memory (RAM) of 16.0 GB. Concerning PSL and EPSL models, the computational time was found to be less than 1 second which is probably due to their simple multinomial logit structure. The small number of considered alternatives per observation may be a further reason for the simplicity of their calculation. However, there is a large difference in computational costs between these models and the IAL model. Although the number of feasible alternatives for the IAL model is limited to 9, the high computational cost might be related to the fact that every possible non-empty subset ($2^9 - 1 = 511$ subsets) must be considered for every observation.

Implementing a NL structure reduces the calculation time substantially, compared to the IAL model, by providing a more realistic structure to represent drivers' route choice behaviour. It is behaviourally not realistic and computationally not feasible to assume that drivers consider 511 non-empty subsets in order to make a choice from a set of 9 nine alternatives. However, the increase in computational time compared to the PSL and EPSL models might be explained by the more complex structure of NL compared to MNL, and the greater number of alternatives considered for the NL model (82 alternatives per observation compared to 20 alternatives for PSL and EPSL models). Expectedly, the largest estimation time is recorded for the LK model. The normally distributed portion of the disturbance requiring large number of draws in model estimation, leads to a computationally demanding model (Ben-Akiva et al., 2001; Walker et al., 2004).

Table 4.4: Model estimates for taxi trips between Montreal and Laval

	PSL		**EPSL**		**IAL**		**LK**		**NL**	
Parameters	*Est.*[a]	*tt.*[b]	*Est.*	*tt.*	*Est.*	*tt.*	*Est.*	*tt.*	*Est.*	*tt.*
$\beta_{Mtl\_Len}{}^{c}$	-0.0009	-10.9	-0.0048	-4.5	-0.0177	-8.4	-0.0012	-5.7	-0.0016	-9.2
$\beta_{Lvl\_Len}{}^{c}$	-0.0002	-3.1	-0.0037	-5.1	-0.0006	-8.8	-0.0003	-2.2	-0.0015	-7.0
$\beta_{Hgw\_Len}{}^{c}$	0.0003	2.4	0.0028	4.8	0.00776	9.5	0.0005	2.7	0.0008	4.3
$\beta_{Seg\_Len}{}^{c}$	0.1070	6.2	0.1910	6.8	-0.197	-8.3	0.1020	11.6	0.0455	6.2
β_{PS}	4.1	6.2								
β_{EPS}			3.57	4.5			2.76	11.5	0.887	3.4
ASC_Brg2					33.6	11.2				
ASC_Brg3					4	-				
ASC_Brg4					18.3	13.2				
ASC_Brg5					-18.8	-4.5				
ASC_Brg6					25	-				
Mu	0.66	7.1	0.38	6.9	1	-	0.65	7.4	1	-
σ_{Brg1}							0.48	2.7		
σ_{Brg2}							-1.12	-2.8		
σ_{Brg3}							-1.81	-4.8		
σ_{Brg4}							0.96	2.9		
σ_{Brg5}							4.35	2.3		
σ_{Brg6}							1.76	1.5		
Mu_Brg2									3.78	17.1
Mu_Brg3									1.08	4.9
Mu_Brg4									3.73	10.6
Mu_Brg5									1.52	6.1
Mu_Brg6									1.42	5.1
Initial LL	-2079.8		-2180.6		-1151.2		-5201.0		-2404.7	
Final LL	-623.9		-606.7		-341.4		-747.7		-336.8	
Rho-square	0.697		0.719		0.697		0.854		0.855	
Est. Time (s)	< 1		< 1		514		2417		123	

[a] Estimated value
[b] Robust t-test
[c] Length has been measured in meters

4.5.2 Validation

In order to further compare these models, their ability to predict should also be evaluated in the final stage. An out of sample validation has been performed to evaluate the prediction capability of the estimated models. Twenty percent of the observations (109 trips), which have not been used for model estimation, have been randomly sampled for this purpose. The validation has been performed based on models' abilities to correctly predict:

i. The chosen bridge,
ii. The chosen route, and
iii. The total overlapping percentage with chosen alternatives (coverage rate).

The aforementioned three indicators have been calculated and results are illustrated in Figure 4.4. Part A of Figure 4.4 compares models' performances in terms of correctly predicting the taken route. It clearly shows that LK and NL perform better than the three other models with a prediction rate of 72.5 % and 73.4 % compared to 51.4 %, 53.2 % and 55 % for the PSL, EPSL and IAL models, respectively. The ability to correctly predict the chosen bridge is compared in part B of Figure 4.4. Similarly, LK and NL outperform PSL and IAL with prediction rates of 91.8 and 92.7 % compared to 89 % and 55 %, respectively.

The performance of EPSL is roughly similar to the anchor-based nested formulations with a small difference of around 1.0 %, which may be attributed to simulation errors. In order to understand how closely each model has predicted the chosen route, we have calculated the length percentage that has been correctly predicted by each model. This coverage rate is compared in part C of Figure 4.4, and clearly demonstrates the superiority of anchor-based nested formulations. Since IAL model is an anchor-based model, it was expected to perform better in bridge choice prediction with respect to studied route-based models. The poor performance of this model, in this study, in terms of both estimation and prediction might be attributed to the fact that a fixed utility function has been used to compare the four models. Using more explanatory variables describing bridge characteristics in Eq. (4.17) might have improved both the model's fit over data and its prediction abilities.

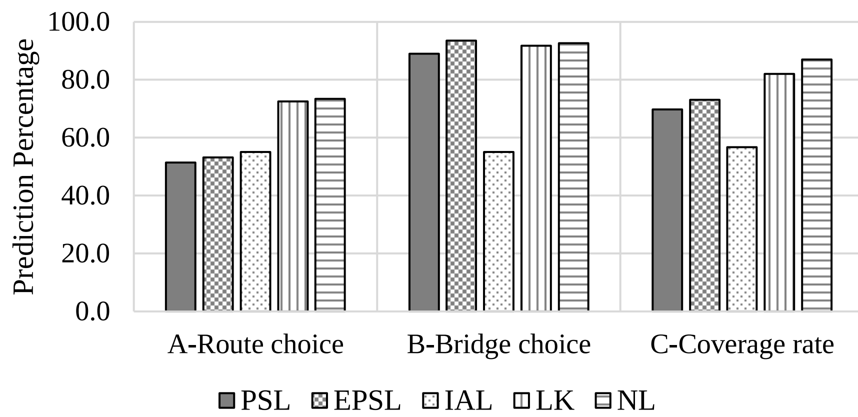


Figure 4.4: Comparison of models' prediction abilities

Based on the abovementioned results, we conclude that the proposed anchor-based nested approaches outperform route-based as well as anchor based models. This conclusion is in line with findings reported by (Habib et al., 2013; Kazagli & Bierlaire, 2015; Manley, Addison, et al., 2015a; Manley, Orr, et al., 2015; Prato & Bekhor, 2007a) who claimed that anchor points have an important effect on individuals' route selection behaviour. However, the important aspect of this study, which is the estimation of the comprehensive effect of both route-level attributes and anchor points, clearly demonstrates that considering the role of bridges as anchor points in conjunction with route-level attributes, for trips between Montreal and Laval, enhances both the estimation and prediction abilities of the model.

4.6 Conclusions

In this paper we have explored the application of a nested structure to improve the behavioural aspect of route choice modelling by incorporating the effect of space hierarchy in drivers' decision making process. The anchor-based nested approaches, proposed in this paper, attempt to improve the behavioural aspect of route choice modelling by incorporating the effects of anchor points and route level attributes at the same time. As previous studies have shown (Habib et al., 2013; Sun et al., 2014; Woo et al., 2015), in riverside cities such as Montreal, route choice decisions are highly influenced by their respective bridge choices. Moreover, routes crossing a same bridge share unobserved components such as safety, scenery, driving comfort, etc., which is mainly because they share the same network and geographical characteristics. In this study, we capture this reciprocal effect through a nested structure. GPS data provided by a taxi company in Montreal has been used to study trips made between the island cities of Montreal and Laval. For these trips, drivers have to choose between a maximum of nine bridges, providing plausible route alternatives between these cities. These bridges face recurrent congestion and play an important role in drivers' route decisions due to their high travel time variability. To address the role of bridges as anchor points, a discrete choice utility maximization framework has been adopted. First, a NL formulation has been proposed, in which upper nests represent bridges and lower nests consist of route alternatives crossing respective bridges. Second, a nested LK with a factor analytic structure is specified. The EPS factor has been added to the deterministic part of the utility functions of these models to account for physical overlap among routes crossing the same bridge. The unobserved similarities among these routes are captured through the nested structure and the factor analytic structure in NL and LK models, respectively.

To evaluate the performance of the proposed anchor-nested formulations, they are compared to the recent route- and anchor-based models, namely PSL, EPSL and IAL models. For the sake of simplicity in estimation and comparison, four most important variables have been selected to define a common utility function between models. Findings revealed that the nested structures provided better model fits and underscored the importance of considering the comprehensive effect of anchor points and route level attributes in route choice decisions. Results have expectedly illustrated that taxi drivers are more likely to drive on highways and tend to decrease their travelled distance. It has also been found that they prefer to avoid intersections and tend to drive on routes with a higher average segment length. The predictive ability of these models has also been compared by an out-of-sample validation approach. Three indicators have been used for this purpose, namely the number of correctly predicted routes, the number of correctly predicted bridges, and the overlap percentage between the predicted and chosen alternatives. The overall results suggest that LK and NL outperform the other three models in the validation step.

In short, incorporating the effects of anchor points in a nested structure has the following advantages over the previously studied anchor based model: (1) the nested structure improves the behavioural aspect of decision making process. In the conventional anchor based model (IAL), a probability is assigned to every subset of the universal choice set, and the conditional probability of an alternative being chosen implies that the decision maker has considered every possible combinations of alternatives as his final choice set, which is behaviourally unrealistic, (2) LK and NL models are easily manageable and practical, even by considering a large number of alternatives

(McFadden, 1978). IAL incorporates a full probabilistic choice set generation approach, which is inapplicable in route choice modelling (Prato, 2009b). For instance, considering a small dataset of 10 alternatives, a selection probability has to be calculated for every 1023 non-empty subsets of alternatives, which is very time-consuming and impractical, and (3) the nested structure allows the consideration of multiple anchor points and their effects on route choice decisions. Also, the inclusion of multiple landmarks and anchor-points, and the consideration of several forms of heterogeneity, such as decision makers' taste variations, are much more manageable and can be accommodated more easily in LK than in the NL mode. This is due to the flexible structure of the error term, which can approximate almost any desirable error structure (Walker et al., 2004).

This study contributes to the existing literature in two ways. First, it improves the behavioural, theoretical, and practical aspects of anchor-based route choice models by capturing the effects of both anchor points and route level attributes within a nested choice model framework, and clearly underscores the importance of considering the effects of anchor points in conjunction with route-level attributes. Second, a large real-world road network, consisting of over 40,000 nodes and 19,000 links, has been studied and a MH algorithm has been adopted to generate a set of considered alternatives. To the best of the authors' knowledge, the largest network previously tested on this algorithm was composed of about 8,000 nodes and 17,000 links (Flötteröd & Bierlaire, 2013). The major advantage of MH sampling algorithm over conventional methods (e.g. link labelling, link elimination, etc.) is that it provides researchers with path sampling probabilities, so that model estimates based on these sets are not biased. It is noted that in route-based models and most of the link-based formulations, the consideration set is commonly generated using shortest-path algorithms with some pre-defined impedance function, which do not provide path sampling probabilities, and hence do not account for the correlation between sampled and non-sampled paths, resulting in biased estimates.

It is worth mentioning that since this paper is based on a dataset covering a fraction of taxi fleets operating in Montreal, results may not be directly transferable to other car drivers in Montreal or any other similar regions in the world. Additional datasets from different contexts and population segments are needed to provide more insights on this subject. However, it reveals the undeniable effect of anchor points on the decision of the whole path, and provides valuable insights regarding drivers' route selection behaviour. As the core of traffic assignment methods, a more realistic route choice model improves the travel demand assessment on the road network. An application instance of the proposed models would be their utilities in predicting drivers' behaviour under hypothetical situations, and the way drivers react to different policies and changes. For example, the effect of a temporary lane closure on one of the bridges can be assessed on other bridges and route segments.

For future works, it would be interesting to investigate the spatial and temporal transferability of the proposed structure for different datasets on similar case studies. Also, more interesting structures such as multilevel nested models can be estimated to explore the effects of multiple anchor points on route choice decisions. In this work, we have neglected the effect of shared segments between routes crossing different bridges (alternatives in different nests), to accommodate the IIA property of the NL model. The inclusion of a correction factor accounting for this shared similarity might improve estimation results.

Route choice is also influenced by travel time and congestion. A shortcoming of this study is that travel time related attributes have been neglected in this study due to data availability issues. In order to improve models' estimation and prediction abilities, it is recommended to consider travel time related attributes on bridges and route segments. Furthermore, including physical characteristics of bridges can also be interesting and can provide useful insights on their effects on the attractiveness of an alternative. It is expected that the inclusion of these factors will enhance models' estimation and prediction abilities. Another appealing aspect to investigate would be the incorporation of some socio-demographic and behavioural factors which were not available in our dataset. Also, incorporating the role of dynamic information, knowledge, and level of experience would add an interesting aspect to this modelling process. A recent study by Vitetta (2016) explores a new realm of route choice models called Quantum Utility Model (QUM), which captures the effect of intermediate (during the trip) decisions, where decision makers are uncertain about their final choices. A comparison study between the proposed approach in this study and the study by Vitetta (2016) might also be interesting as a future expansion of this work. Another interesting area would be the comparison of the results with the Recursive Logit (RL) modeling framework.

Acknowledgement

The authors are grateful to Gunnar Flötteröd for his invaluable help with the implementation of MH algorithm. We also acknowledge collaborators from Taxi Diamond who provided access to data for research purpose, as well as the anonymous reviewers for their helpful comments and suggestions. We would also thank the Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) for providing the possibility of using their cluster computer for our research funded by the Natural Sciences and Engineering Council of Canada (NSERC) Research Tools and Instruments grant.

References

Arifin, Z.N.: Route choice modeling based on GPS tracking data. Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 20400, 2012 (2012)

Azevedo, J., Costa, M.E.O.S., Madeira, J.J.E.S., Martins, E.Q.V.: An algorithm for the ranking of shortest paths. European Journal of Operational Research 69(1), 97-106 (1993)

Bekhor, S., Ben-Akiva, M., Scott Ramming, M.: Adaptation of logit kernel to route choice situation. Transportation Research Record: Journal of the Transportation Research Board(1805), 78-85 (2002)

Ben-Akiva, M.: Structure of passenger travel demand models. Massachusetts Institute of Technology (1973)

Ben-Akiva, M., Bergman, M., Daly, A.J., Ramaswamy, R.: Modeling inter-urban route choice behaviour. In: Proceedings of the 9th International Symposium on Transportation and Traffic Theory, VNU Press, Utrecht 1984, pp. 299-330

Ben-Akiva, M., Bierlaire, M.: Discrete choice methods and their applications to short term travel decisions. In: Handbook of transportation science. pp. 5-33. Springer, (1999)

Ben-Akiva, M., Bierlaire, M.: Discrete choice models with applications to departure time and route choice. Handbook of transportation science 32 (2003)

Ben-Akiva, M., Bolduc, D., Walker, J.: Specification, identification and estimation of the logit kernel (or continuous mixed logit) model. Department of Civil Engineering Manuscript, MIT (2001)

Bierlaire, M.: BIOGEME: a free package for the estimation of discrete choice models. In: Swiss Transport Research Conference 2003, vol. TRANSP-OR-CONF-2006-048

Bierlaire, M.: BisonBiogeme 2.4: estimating a first model. Series on Biogeme TRANSP-OR 150720 (2015)

Bierlaire, M., Bolduc, D., McFadden, D.: The estimation of generalized extreme value models from choice-based samples. Transportation Research Part B: Methodological 42(4), 381-394 (2008). doi:http://dx.doi.org/10.1016/j.trb.2007.09.003

Bierlaire, M., Fetiarison, M.: Estimation of discrete choice models: extending BIOGEME. In: Swiss Transport Research Conference (STRC) 2009

Bierlaire, M., Frejinger, E.: Route choice models with subpath components. In: Swiss Transportation Research Conference 2005, vol. TRANSP-OR-CONF-2006-032

Bierlaire, M., Frejinger, E.: Route choice modeling with network-free data. Transportation Research Part C: Emerging Technologies 16(2), 187-198 (2008)

Bolduc, D., Ben-Akiva, M.: A multinomial probit formulation for large choice sets. In: Proceedings of the 6th International Conference on Travel Behaviour 1991, pp. 243-258

Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A.: A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks. In: Proceedings of the 13th International Symposium on Transportation and Traffic Theory 1996, pp. 697-711. Pergamon Oxford, NY, USA

Cascetta, E., Russo, F., Viola, F.A., Vitetta, A.: A model of route perception in urban road networks. Transportation Research Part B: Methodological 36(7), 577-592 (2002)

Communauté métropolitaine de montréal, C.: An Attractive, Competitive and Sustainable Greater Montreal. In. Library and Archives Canada Montréal, Québec, (2012)

Connors, R.D., Sumalee, A.: A network equilibrium model with travellers' perception of stochastic travel times. Transportation Research Part B: Methodological 43(6), 614-624 (2009).

Couclelis, H., Golledge, R.G., Gale, N., Tobler, W.: Exploring the anchor-point hypothesis of spatial cognition. Journal of Environmental Psychology 7(2), 99-122 (1987)

de la Barra, T., Perez, B., Anez, J.: Multidimensional path search and assignment. In: PTRC Summer Annual Meeting, 21st, 1993, University of Manchester, United Kingdom 1993

Dhakar, N., Srinivasan, S.: Route choice modeling using GPS-based travel surveys. Transportation Research Record: Journal of the Transportation Research Board(2413), 65-73 (2014)

Dougherty, M.: A review of neural networks applied to transport. Transportation Research Part C: Emerging Technologies 3(4), 247-260 (1995).

Duan, Z., Wei, Y.: Revealing Taxi Driver Route Choice Characteristics Based on GPS Data. CICTP 2014 Safe, Smart, and Sustainable Multimodal Transportation Systems, 565-573 (2014)

Elgar, I., Farooq, B., Miller, E.: Modeling location decisions of office firms: Introducing anchor points and constructing choice sets in the model system. Transportation Research Record: Journal of the Transportation Research Board(2133), 56-63 (2009)

Elgar, I., Farooq, B., Miller, E.J.: Simulations of firm location decisions: Replicating office location choices in the Greater Toronto Area. Journal of Choice Modelling 17, 39-51 (2015).

Flötteröd, G., Bierlaire, M.: Metropolis–Hastings sampling of paths. Transportation Research Part B: Methodological 48, 53-66 (2013)

Foo, P., Warren, W.H., Duchon, A., Tarr, M.J.: Do humans integrate routes into a cognitive map? Map-versus landmark-based navigation of novel shortcuts. Journal of Experimental Psychology: Learning, Memory, and Cognition 31(2), 195 (2005)

Fosgerau, M., Frejinger, E., Karlstrom, A.: A link based network route choice model with unrestricted choice set. Transportation Research Part B: Methodological 56, 70-80 (2013).

Frejinger, E., Bierlaire, M.: On path generation algorithms for route choice models. SH Daly and A.(eds.) Choice Modelling: The State-of-the-Art and the State-of-Practice, 307-315 (2010)

Frejinger, E., Bierlaire, M., Ben-Akiva, M.: Sampling of alternatives for route choice modeling. Transportation Research Part B: Methodological 43(10), 984-994 (2009)

Gao, S., Frejinger, E., Ben-Akiva, M.: Adaptive route choices in risky traffic networks: A prospect theory approach. Transportation Research Part C: Emerging Technologies 18(5), 727-740 (2010). doi:http://dx.doi.org/10.1016/j.trc.2009.08.001

Golledge, R.G., Smith, T.R., Pellegrino, J.W., Doherty, S., Marshall, S.P.: A conceptual model and empirical analysis of children's acquisition of spatial knowledge. Journal of Environmental Psychology 5(2), 125-152 (1985)

Guevara, C.A., Ben-Akiva, M.E.: Sampling of alternatives in Multivariate Extreme Value (MEV) models. Transportation Research Part B: Methodological 48, 31-52 (2013)

Habib, K.N., Morency, C., Trépanier, M., Salem, S.: Application of an independent availability logit model (IAL) for route choice modelling: Considering bridge choice as a key determinant of selected routes for commuting in Montreal. Journal of choice modelling 9, 14-26 (2013)

Henn, V.: Route Choice Making Under Uncertainty: a Fuzzy Logic Based Approach. In: Verdegay, J.-L. (ed.) Fuzzy Sets Based Heuristics for Optimization, vol. 126. Studies in Fuzziness and Soft Computing, pp. 277-292. Springer Berlin Heidelberg, (2003) Hess, S., Daly, A.J.: Choice Modelling: The State-of-the-Art and the State-of-Practice: Proceedings from the Inaugural International Choice Modelling Conference: The State-of-the-art and the State-of-practice: Proceedings from the Inaugural International Choice Modelling Conference. Emerald Group Publishing, (2010)

Hess, S., Quddus, M., Rieser-Schüssler, N., Daly, A.: Developing advanced route choice models for heavy goods vehicles using GPS data. Transportation Research Part E: Logistics and Transportation Review 77, 29-44 (2015)

Hirtle, S.C., Jonides, J.: Evidence of hierarchies in cognitive maps. Memory & cognition 13(3), 208-217 (1985)

Holding, C.S.: Further evidence for the hierarchical representation of spatial information. Journal of Environmental Psychology 14(2), 137-147 (1994).

Kahneman, D., Tversky, A.: Prospect theory: An analysis of decision under risk. Econometrica: Journal of the Econometric Society, 263-291 (1979)

Kaplan, S., Prato, C.G.: Closing the gap between behaviour and models in route choice: The role of spatiotemporal constraints and latent traits in choice set formation. Transportation Research Part F: Traffic Psychology and Behaviour 15(1), 9-24 (2012).

Kazagli, E., Bierlaire, M.: A Route Choice Model based on Mental Repre-sentations. (2015)

Kim, J., Sung, S., Namgung, M., Jang, Y.: Development of dynamic route choice behavioral applied intelligent system theory. In: Proceedings of the Eastern Asia Society for Transportation Studies 2005, pp. 1615-1630

Lai, X., Bierlaire, M.: Specification of the cross-nested logit model with sampling of alternatives for route choice models. Transportation Research Part B: Methodological 80, 220-234 (2015)

Luisa De Maio, M., Vitetta, A.: Route choice on road transport system: A fuzzy approach. Journal of Intelligent & Fuzzy Systems 28(5) (2015)

Lynch, K.: The image of the city, vol. 11. MIT press, (1960)

Manley, E., Addison, J., Cheng, T.: Shortest path or anchor-based route choice: a large-scale empirical analysis of minicab routing in London. Journal of Transport Geography 43, 123-139 (2015a)

Manley, E., Orr, S., Cheng, T.: A heuristic model of bounded route choice in urban areas. Transportation Research Part C: Emerging Technologies 56, 195-209 (2015b)

Manski, C.F.: The structure of random utility models. Theory and decision 8(3), 229-254 (1977)

McFadden, D.: Modelling the choice of residential location. Institute of Transportation Studies, University of California, (1978)

McFadden, D., Train, K.: Mixed MNL models for discrete response. Journal of applied Econometrics, 447-470 (2000)

McFadden, D.L.: Econometric analysis of qualitative response models. Handbook of econometrics 2, 1395-1457 (1984)

Murat, Y.S., Uludag, N.: Route choice modelling in urban transportation networks using fuzzy logic and logistic regression methods. Journal of Scientific and Industrial research 67(1), 19 (2008)

Prato, C., Bekhor, S.: Applying branch-and-bound technique to route choice set generation. Transportation Research Record: Journal of the Transportation Research Board(1985), 19-28 (2006)

Prato, C., Bekhor, S.: Modeling route choice behavior: how relevant is the composition of choice set? Transportation Research Record: Journal of the Transportation Research Board (2007)

Prato, C.G.: Route choice modeling: past, present and future research directions. Journal of Choice Modelling 2(1), 65-100 (2009). doi:http://dx.doi.org/10.1016/S1755-5345(13)70005-8

Prato, C.G.: Expanding the applicability of random regret minimization for route choice analysis. Transportation 41(2), 351-375 (2014)

Prato, C.G., Bekhor, S., Pronello, C.: Latent variables and route choice behavior. Transportation 39(2), 299-319 (2012)

Quattrone, A., Vitetta, A.: Random and fuzzy utility models for road route choice. Transportation Research Part E: Logistics and Transportation Review 47(6), 1126-1139 (2011).

Ramming, M.S.: Network knowledge and route choice. MIT (2002)

Sun, H., Wu, J., Ma, D., Long, J.: Spatial distribution complexities of traffic congestion and bottlenecks in different network topologies. Applied Mathematical Modelling 38(2), 496-505 (2014). doi:http://dx.doi.org/10.1016/j.apm.2013.06.027

Swait, J., Ben-Akiva, M.: Incorporating random constraints in discrete models of choice set generation. Transportation Research Part B: Methodological 21(2), 91-102 (1987)

Train, K.E.: Discrete choice methods with simulation. Cambridge university press, (2009)

Tversky, A., Kahneman, D.: Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and uncertainty 5(4), 297-323 (1992)

Vitetta, A.: A quantum utility model for route choice in transport systems. Travel Behaviour and Society 3, 29-37 (2016). doi:http://dx.doi.org/10.1016/j.tbs.2015.07.003

Vovsha, P., Bekhor, S.: Link-nested logit model of route choice: overcoming route overlapping problem. Transportation Research Record: Journal of the Transportation Research Board(1645), 133-142 (1998)

Walker, J., Ben-Akiva, M., Bolduc, D.: Identification of the logit kernel (or mixed logit) model. In: 10th International Conference on Travel Behavior Research, Lucerne, Switzerland, August 2004

White, C.E., Bernstein, D., Kornhauser, A.L.: Some map matching algorithms for personal navigation assistants. Transportation Research Part C: Emerging Technologies 8(1), 91-108 (2000)

Wiener, J.M., Mallot, H.A.: 'Fine-to-coarse'route planning and navigation in regionalized environments. Spatial cognition and computation 3(4), 331-358 (2003)

Woo, C.K., Cheng, Y.S., Li, R., Shiu, A., Ho, S.T., Horowitz, I.: Can Hong Kong price-manage its cross-harbor-tunnel congestion? Transportation Research Part A: Policy and Practice 82, 94-109 (2015). doi:http://dx.doi.org/10.1016/j.tra.2015.09.002

Xu, H., Zhou, J., Xu, W.: A decision-making rule for modeling travelers' route choice behavior based on cumulative prospect theory. Transportation Research Part C: Emerging Technologies 19(2), 218-228 (2011). doi:http://dx.doi.org/10.1016/j.trc.2010.05.009

Zhou, J., Golledge, R.: A three-step general map matching method in the GIS environment: Travel/transportation study perspective. International Journal of Geographical Information System 8(3), 243-260 (2006)

CHAPTER 5 ARTICLE 2: CLASSIFYING BEHAVIOURAL DYNAMICS OF TAXI DRIVERS OPERATING STRATEGIES USING LONGITUDINAL ROUTE CHOICE DATA

Hamzeh Alizadeh[1*], Bilal Farooq[2], Catherine Morency[1], Nicolas Saunier[1]

Under review (revision requested) in *Transportation Research Part F: Traffic Psychology and Behaviour*

Submission Date: 30 / November / 2017

[1] Department of Civil, Geological and Mining Engineering, Polytechnique Montréal, C.P. 6079, succ. Centre-Ville Montréal, Canada

[2] Department of Civil Engineering, Ryerson University, 350 Victoria Street, Toronto, Canada

[*] Corresponding author

Abstract

Traditionally, taxi service has provided an important mean of transportation in large cities, acting mainly as a temporal and spatial complement to the existing public transport. Taxi drivers are more familiar with the road network compared to normal users. Due to their more extensive driving experience, they develop various habits and behaviours, which breed different types of operating strategies. We aim to classify these strategies based on revealed preference route choice decisions. We have studied a longitudinal GPS dataset, tracking 1,746 taxi drivers over one year, and performed a Principal Component Analysis followed by a Hierarchical Agglomerative Clustering to extract behavioural clusters. Four major types of route choice behaviour are observed. These clusters show significant variations based on the time of day and the travelled distance, and are labelled: *"Short trips night drivers"*, *"Long trips night drivers"*, *"Short trips day drivers"*, and *"Long trips day drivers"*. The understanding of these patterns is important to transportation planners to encourage a shared economy by proposing fair policies for the incumbent of taxi industry and other transportation companies. Furthermore, the inclusion of different classes of drivers in route choice models would improve their behavioural aspects as well as their estimation and prediction abilities.

Keywords: Taxi, Operating strategies, Route choice behaviour, Longitudinal GPS data, Principal component analysis, Hierarchical clustering

5.1 Introduction

Taxi service provides a prompt, rapid, comfortable, convenient, twenty-four seven service, which makes it an important mean of transportation in large cities. This alternative transportation service acts mainly as a temporal and spatial complement to the existing public transport, i.e. bus, metro, etc. (Shaaban & Kim, 2016). Taxi drivers are usually more familiar with the road network and its travel time variations, compared to normal users. Due to their extensive driving experience, they develop different habits and behaviours, which breed different types of operating strategies (Eagle & Pentland, 2009; Kang & Qin, 2016). The understanding of these strategies helps to better comprehend urban traffic dynamics and is very important to the city and transportation planners (Manley, Addison, et al., 2015a).

These strategies can be observed and characterized through drivers' route choices, which form the collective behavioural patterns of taxi movement. The complex process of route selection is captured through route choice models, which are used to estimate and predict the probability of a certain route being chosen between a given OD pair. Route choices are highly dependent on individuals' characteristics such as having different preferences, experiences, information levels, and attitudes. Our hypothesis is that these factors are correlated and can be classified to represent various types of route choice behaviours, hence, operating strategies.

In this study, we aim to extend the understanding of taxi drivers' operating strategies by classifying them into separate categories, based on their observed route choice decisions, using a longitudinal GPS dataset. Route choice studies have been mostly based on traditional cross-sectional analysis of drivers' behaviour, hence, data for a single day has usually been interpreted to derive the most important attributes affecting drivers' decision-making patterns. In recent years, with the dissemination of GPS technologies and the progressive utilization of in-vehicle or on-person GPS devices, researchers are supplied with longitudinal choice data over an extended period that provides better insights on drivers' route choice patterns and its inert heterogeneity.

This research relies on a longitudinal GPS dataset, tracking 1,746 taxi drivers making more than 22,000 trips over a period of one year. In order to extract behavioural clusters, a Principal Component Analysis (PCA) is first used to reduce dataset's dimensions, by forming uncorrelated variables while retaining as much information as possible. Then a Hierarchical Agglomerative Clustering (HAC) is performed to classify drivers' behaviours.

This work distinguishes itself from the existing literature based on two original contributions. First, to the extent of our knowledge, drivers' operating strategies have never been stratified using drivers' actual route choices. The incorporation of different driver categories can improve the estimation and prediction accuracy of route choice models, and can be used in defining different functional forms for traffic assignment models (Parkany et al., 2006a; Ramaekers et al., 2013; Tawfik et al., 2010c). Second, route choice preferences have never been looked at through longitudinal GPS data over a one-year period. Previous datasets were not large or detailed enough to allow this type of stratification. Studies focusing on driving patterns over an extended period appear to be minimal, which might be mostly due to the challenges of data collection and analysis.

The rest of this paper is organized as follows. First, we overview the state-of-the art on road users behavioural classification, and factors affecting drivers' route choice decisions. Second, we present our studied regional context and describe our dataset. Then, we discuss the process of deriving indicators and our clustering methodology. Finally, we present some descriptive analysis of the studied trips, discuss the clustering results and underline the findings of this study.

5.2 State-of-the-art

Classification, i.e. clustering, is a machine learning technique of grouping similar observations into separate homogeneous groups (clusters), with respect to some predefined measures of similarity (Hair, 2009). In transportation related studies, behavioural classifications have been conducted on various segments of the population, such as pedestrians (Okamoto et al., 2011), bike users (Damant-Sirois et al., 2014; Dill & McNeil, 2013; Geller, 2009; Kroesen & Handy, 2014), bike-sharing (Reinoso & Farooq, 2015), and carsharing members (Morency et al., 2011).

Kroesen and Handy (2014) studied the causality relation between non-work-related trips and commuters, and classified cyclists into four different clusters, namely *Non-cyclists*, *Non-work cyclists*, *All-around cyclists*, and *Commuter cyclists*. Another major work on cyclists classification is the work by Geller (2009), in which bike users are classified into four categories: *Strong and Fearless*, *Enthused and Confident*, *Interested but concerned*, and *No Way No How*. The reader is referred to Dill and McNeil (2013) for more reflections on this work. A recent work by Damant-Sirois et al. (2014) classified Montreal cyclists, based on the intensity of bicycle usage, into four groups, namely *dedicated*, *path-using*, *fair-weather utilitarians*, and *leisure* cyclists. Furthermore, on the bike sharing side, Reinoso and Farooq (2015) found two types of users: *Commuters* and *Recreational users*. In a study by Morency, Trepanier, and Agard (2011), behaviours of car sharing members have been classified, based on two main indicators, namely the number of transactions and distance travelled, each resulting in two main clusters. In addition, weekly patterns of transactions have been used to estimate the regularity of active weeks of members, showing that more than 60 % of members have similar weeks of usage.

Classification of driving behaviours has been mostly cited in traffic safety studies to distinguish between conflict and collision leading behaviours, risk-taking versus safe drivers and experienced versus novice drivers (Lucidi et al., 2010; Marengo, Settanni, & Vidotto, 2012; Saunier, Mourji, & Agard, 2011; Ulleberg, 2001), as well as in car-following research, to classify drivers according to their driving behaviours (Higgs & Abbas, 2013). For instance, a classification of driving patterns has been proposed by Jensen (1999), which divides drivers into three categories based on their attitudes towards the environment: *Passionate*, *Every day*, and *Leisure time* drivers.

Drivers' perceptions and experiences have been studied by Tawfik et al. (2010a), using a driving simulator and two initial and final questionnaires. They found observable differences between drivers' route choice behaviours, and categorized drivers based on their learning skills into four different categories. Tawfik, Szarka, House, and Rakha (2011a) have demonstrated that the inclusion of those learning clusters would improve disaggregated route choice models. In a further study by Liu, Andris, and Ratti (2010), trip characteristics and incomes of taxi drivers have been studied to classify them into *Top Drivers* and *Ordinary Drivers*. They have illustrated that top drivers determine their operational zones based on a trade-off between travel demand and traffic conditions and prefer to avoid highly congested areas and operate in districts where travel demand is higher. On the opposite side, ordinary drivers ignore the effect of traffic conditions and operate in the central business district in rush hours, most probably to profit from the higher travel demand. They also argue that these two classes of taxi drivers show significant differences with respect to trip distances, trip times, and the ratio of observed path lengths and times over the respective time and distance based-shortest paths. Taxi drivers have been the subject of another study by Kang and Qin (2016), in which their spatial operation behaviours have been studied using digital traces of 6000 taxis over a period of one month. They adopt a non-negative matrix factorization method to better understand patterns of taxi demand-and-supply and uncover self-organized spatial operation patterns of taxi drivers.

Although these studies, among others, show that different categories of road users are observable, an explicit classification of taxi drivers' operating strategies based on their actual route choices is missing from the literature. Due to the availability of very large and rich datasets from taxi companies, that have detailed GPS trajectories of their fleets for a long duration of time, it has become possible to address this critical research need.

It has been well established in the existing literature that drivers tend to choose their routes in a way that minimizes the incurred cost in a given time period. This cost function is unique for each driver and situation, and can be attributed to several factors, such as different levels of information, different capacities to process them, and different computational and prediction abilities (Ben-Akiva et al., 1991). Bovy and Stern (1990b) argued that travel behaviour is influenced by four categories of attributes, namely *Physical*, *Socio-demographic*, *Normative*, and *Personal* attributes. *Physical* components include network characteristics and travel possibilities. Factors such as household characteristics, age and gender are part of the *Socio-demographic* components, and *Normative* attributes encompass norms and values derived from the society surrounding the traveller. These three categories along with *Personal* factors, which are basically the preferences and attitudes of the decision maker, affect his/her travel behaviours. Furthermore, they argue that the mental process of route choice is influenced by personal preferences, characteristics, and attitudes of the decision maker, which are in turn affected by his/her acquired knowledge of the network, level of information, and experience.

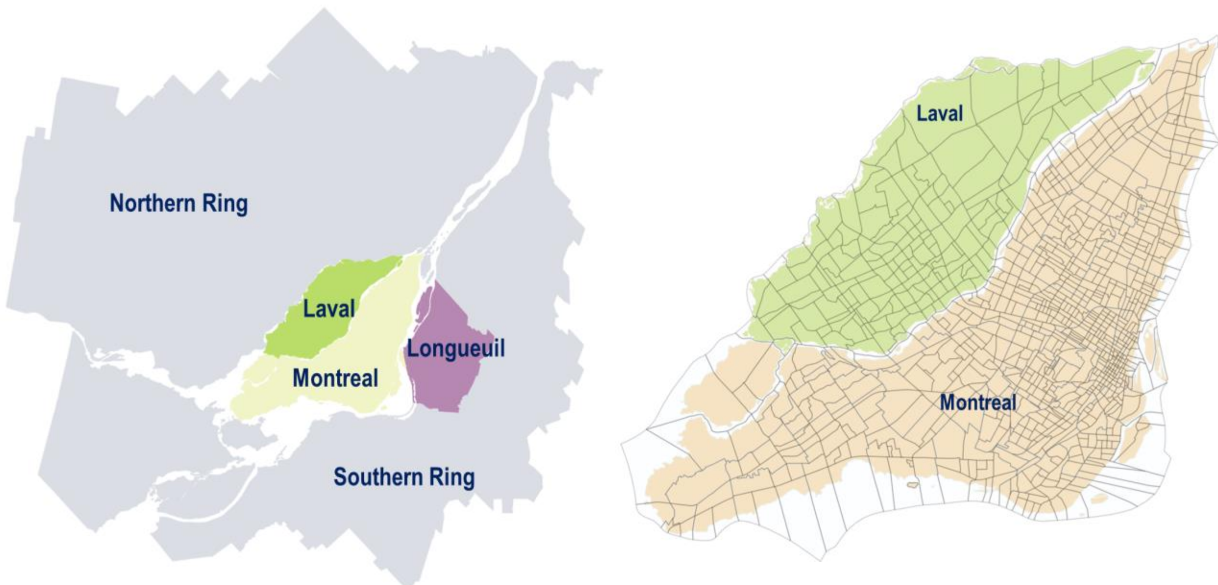
Each route choice experiment can affect drivers' personal preferences by its satisfactory or unsatisfactory outcome (Dia, 2002). It is argued that the repetition of satisfactory results can become a stable preference and drivers tend to repeat same choices, forming various commute patterns (Ben-Akiva et al., 1991). For taxi drivers, these commute patterns can be interpreted as their operating strategies. In this study, we aim to classify these strategies based on revealed preference route choice decisions.

5.3 Context and data description

5.3.1 Regional Context

This study is performed in the context of Greater Montreal Area, depicted in Figure 5.1(a). This Island city is separated from its suburbs by two large rivers; Prairies River and Saint Lawrence

River in the north and south of the city, respectively. In this study, we focus on trips originating in Montreal with a destination in Laval, the largest suburb of Montreal located north of the city, across the Prairies River. These two islands contain roughly 2.3 million inhabitants, and cover a total surface of 632.3 km2 (Communauté métropolitaine de montréal, 2012). Traffic Analysis Zones (TAZ), delineated by Quebec's Ministry of Transport, Sustainable Mobility and Electrification of Transport, are used in this study to determine the characteristics of taxi trips' origin and destination points (Figure 5.1(b)). TAZs are geographical areas, which divide the city into smaller similar areas based on various factors such as population, demography, socioeconomic information, road network, transit access, land use and topography.



(a) Greater Montreal. (b) TAZ of Montreal and Laval.

Figure 5.1: Regional context of the study

5.3.2 GPS Dataset

We investigated GPS traces collected over a period of one year (from January 1st to December 31st, 2015) by a major taxi company operating in Montreal. This taxi company constitutes around 25 % of the Montreal's Island taxi fleet, and its operation is restricted to trips starting or ending in the central part of the island (Lacombe & Morency, 2016; Pele & Morency, 2014). Every taxi is equipped by a data logger and GPS data are collected continuously for operational purposes. For our study, we extracted a subset of the main dataset, consisting of around 750,000 GPS records collected from 1,746 taxi drivers making a total of 22,394 trips. Drivers are associated with a unique

ID so that we can distinguish between trips made by a same car and different drivers. We should mention that personal information on drivers' demographic and socioeconomic characteristics are not available.

In order to explore factors affecting drivers' route choices, we first need to process the GPS dataset to derive the observed routes. Therefore, every record has to be associated with a link in the network, a process called map-matching. This step is crucial, since it determines the accuracy of reconstructed trajectories, and accordingly their derived attributes. Data has been stored in a PostgreSQL database, and the PostGIS spatial extension has been added to support geographical datatypes and queries. To associate GPS records to the road network, a direction-based nearest link point-to-curve map matching algorithm has been adopted, in which every record is matched onto the closest link in the network with respect to its azimuth. To reconstruct the complete trajectory, consecutive GPS points have been connected using a distance-based shortest-path algorithm.

5.3.3 Network Dataset

The road network has been extracted from the OpenStreetMap project in the format of geographical layers (shapefiles). It contains more than 156,000 nodes and 89,000 links. The network has been made routable, through a geospatial extension of PostgreSQL named "pgrouting", in order to enable the calculation of shortest paths. It has also been segmented on an intersection-to-intersection basis; therefore, links are defined to be road segments between two consecutive intersections.

5.4 Methods

5.4.1 Process of Deriving Indicators

The first step towards classifying taxi drivers' operating strategies is to explore a large range of route choices and factors affecting them. These factors include temporal and environmental attributes, drivers' attitudes and preferences, network familiarity, personal demographic and socio-economic characteristics, and route related attributes among others. Since personal characteristics are not observable through GPS traces, the focus of this section is to derive explanatory factors from the GPS dataset to thoroughly describe observed trajectories. Factors derived in this study are classified into five broad categories, namely *Temporal Indicators*, *Route Attributes*, *Driver Characteristics*, *Land Use*, and *Route Similarities*.

Since temporal variations might affect the chosen route, the timestamp information of the origin point of a trip, containing information regarding the date and time of the record, has been used to derive *Temporal Indicators*. Then, in order to characterize each observed route, its physical specifications and similarity level with its respective distance-based shortest path alternative has been evaluated (*Route Attributes*). To derive *Driver Characteristics*, the total number of trips per driver and trips made by the same driver between same TAZ pairs have been inspected and several attributes, pertinent to the regularity of drivers between these TAZ pairs, have been developed. Although GPS data allows us to locate the exact pick up and drop off position of travellers, we have no information regarding their exact trip purposes. In this work, we associated GPS points to their respective TAZ to obtain information regarding the dominant *Land Use* of trips' origin and destination points. Land use data was provided by Quebec's Ministry of Transport, Sustainable Mobility and Electrification of Transport. Four categories of land uses are considered in this study, namely residential, commercial, work/study, and recreational.

To evaluate *Route Similarities* for trips taken by a particular driver between a given pair of TAZ, we have used a measure called Path-Size (PS). This measure has been proposed by Ben-Akiva and Bierlaire (1999a) to account for similarities between routes in logit based discrete choice models. The following formulation has been adopted in this study:

$$PS_{in} = \sum_{a \in \Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j \epsilon \varphi_n} \delta_{aj}} \tag{5.1}$$

where PS_{in} denotes the path-size factor for driver n and route i, L_a and L_i represent the length of link a and route i, Γ_i is the set of road segments in route i, φ_n denotes the observed routes for driver n between the same pair of TAZ, $\sum_{j \epsilon \varphi_n} \delta_{aj}$ indicates the total number of alternatives in φ_n sharing link a (δ_{aj} is the link-path incident binary variable which is 1 if link a is on route i, and 0 otherwise). The upper bound value of this formulation is 1, indicating that observed routes are completely independent and do not share any links. However, smaller values of PS indicate longer overlaps and higher dependencies between trips. Table 5.1 presents a concise description of indicators derived in this study.

Table 5.1: Description of indicators derived to classify drivers’ operating strategies

Indicators	Descriptions
Temporal Indicators	
Peak/Off-peak	Trips starting between 6:00 to 9:00, and 16:00 to 19:00 are considered as peak hour trips, while all other trips are regarded as off-peak trips
Weekday/Weekends	
Day of the Week	
Day/Night	Day trips consist of trips starting between 6:00 to 21:00 and the rest are considered as night trips.
Season	
Route Specifications	
Length	The total travelled distance
Travel time	The timestamp information at origins and destinations have been used to extract the observed travel time
% Highway	Specifies the length percentage of trips made on highways
# Turns / km	The total number of turns per kilometre
# Segments / km	The total number of road segments per kilometre
% Shortest path	% of overlap with the distance-based shortest path
Driver Characteristics	
Total trips	The total number of trips made by each driver
Number of TAZ (#TAZ)	The total number of unique TAZ pairs travelled by each driver
Average number of trip per TAZ pair (TAZ_avg_trip)	The total number of trips, divided by the total number of unique TAZ pairs, between which a driver has travelled. The lower bound of this attribute is one, indicating only one travel per TAZ for a driver. Higher values signify higher propensity of travelling between same TAZ pairs.
TAZ pairs with more than 4 trips (TAZ_4)	The total number of TAZ pairs between which the driver has made more than four trips (4 trips has been selected based on the distribution of trips between same TAZ pairs as an indicator of regularity of drivers).
Maximum same TAZ trips (TAZ_max)	The maximum number of trips made between a given TAZ pair by the same driver. This factor is also interpreted as an indicator of regular trips between same TAZ pairs.
Land Use	
Origin land use	Land use of trip’s origin
Destination land use	Land use of trip’s destination
Route Similarities	
Path-Size (PS)	The average PS factor calculated for all trips made by the same driver, between same TAZ pairs

5.4.2 Cluster Analysis

Clustering is an unsupervised categorization technique that aims to segregate multivariate data sets into more meaningful clusters, according to their main describing attributes. In transportation studies datasets usually have very high dimensions. Multivariate analysis and dimension reduction techniques, such as PCA, are adopted to better interpret the data and to improve the quality of cluster analysis. PCA is an unsupervised dimension reduction technique, which preserve most of the initial information within a smaller numbers of mutually uncorrelated components.

In this study, we adopt a two-step procedure in order to classify driver behaviours. First, in order to reduce collinearity among attributes, a PCA for mixed data is performed (Pagès, 2004). The number of principal components extracted from a PCA analysis depends on the level of correlation between attributes and the amount of variance that can be explained by each principal component. The total number of components is equal to the number of attributes, but only the first few ones are important and are considered in the next step (Poucin, Farooq, & Patterson, 2016). Then, a HAC using Ward's criterion is used to classify drivers' behaviours. In HAC method, clusters are formed hierarchically by merging the two closest clusters at each step. All PCA and clustering analysis are performed in TANAGRA statistical package (Rakotomalala, 2005).

To assess the significance level of attributes in the clustering result, the "Test Value" (TV) criterion has been used. Following formulas are used to calculate TV for continuous (t_c), and discrete (t_d) values for each cluster (Lebart, PIRON, Lebart, Morineau, & Piron, 2000):

$$t_c = \frac{\mu_g - \mu}{\sqrt{\frac{n - n_g}{n - 1} \times \frac{\sigma^2}{n_g})}} \tag{2}$$

$$t_d = \frac{n_{jg} - \frac{n_g \times n_j}{n}}{\sqrt{\frac{n - n_g}{n - 1} \times \left(1 - \frac{n_j}{n}\right) \times \frac{n_g \times n_j}{n}}} \tag{3}$$

where μ and μ_g are attributes' means in the cluster and group, respectively; n and n_g denote the size of the cluster and the group, respectively; σ^2 represents the attribute variance in the cluster; and n_{jg} is the number of observations corresponding to the discrete attribute j in cluster g.

5.5 Results

5.5.1 Trip Characteristics

The spatial distribution of origin and destination points is used to visualize the dispersion of taxi demand for trips between Montreal and Laval. As depicted in the heat map of Figure 5.2, taxi trips mostly originate from downtown Montreal, the airport, major commercial centers, and around train stations. Destination regions with high travel densities are more dispersed in Laval, which is probably due to the higher dispersion of commercial centers and overall lower density. Also, considering that Laval is a suburb of Montreal, the higher segregation between its dense residential areas contributes to the dispersion of high density destination points.

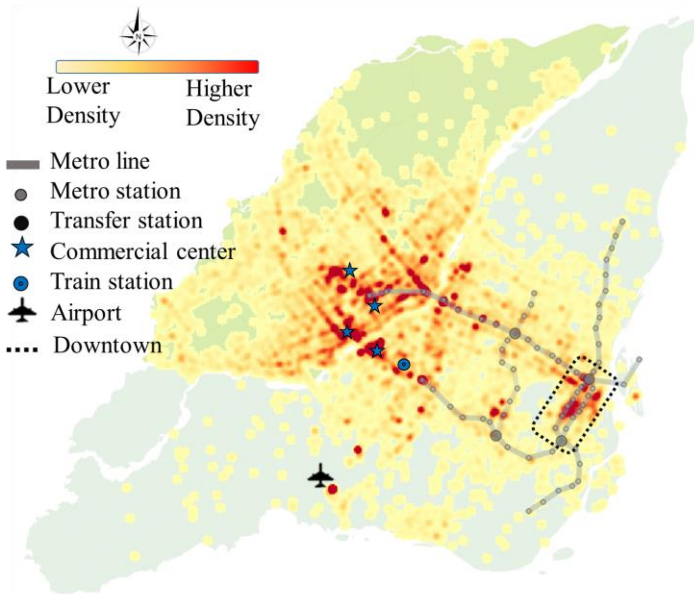
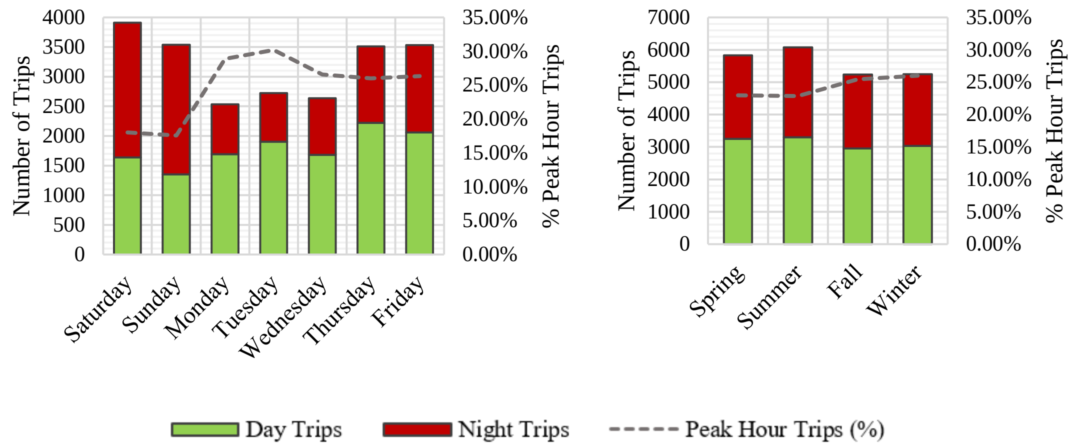
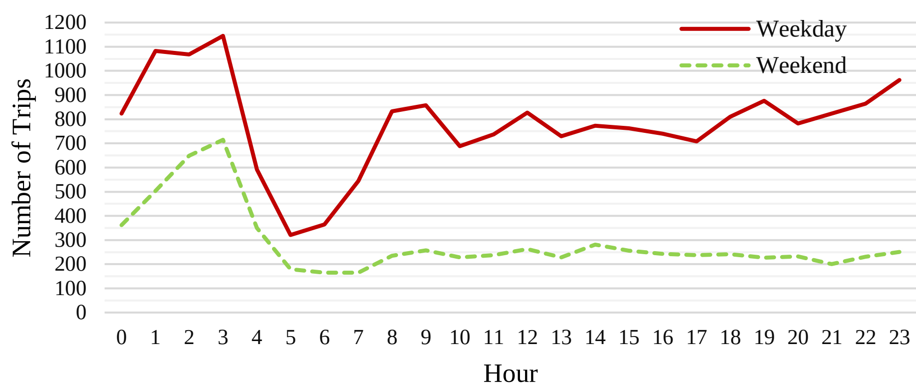


Figure 5.2: Heat map of origin-destination points for trips between Montreal and Laval

A detailed comparison of trip frequencies across different days, the proportion of day trips versus night trips, and the percentage of trips made in peak hours are presented in Figure 5.3(a). A quick look reveals that taxi trips are more frequent during weekends, night trips are more common on Fridays and weekends, and peak hour trips are more frequent during weekdays. The same comparison is illustrated for trips made in different seasons (Figure 5.3(b)). Slightly higher numbers of trips are observed during spring and summer. However, the overall percentage of peak hour trips remains around 24 % throughout the whole year. Overall, around 56 % of trips are day trips and the remaining 44 % are made during nights. These findings are consistent with those reported by Lacombe and Morency (2016), also using data from the most important taxi service provider in Montreal. A closer look at the hourly distribution reveals that the peak demand occurs around 3 AM, drops significantly early in the morning, augments and stays steady over the day, and starts to increase again by the end of the day, around 9 PM (Figure 5.3(c)). The same demand pattern holds for both weekdays and weekends; however, the overall weekday demand is much higher. These findings are in accordance with findings reported by Pele and Morency (2014) using a similar source of data.



(a) Daily variations. (b) Seasonal variations.

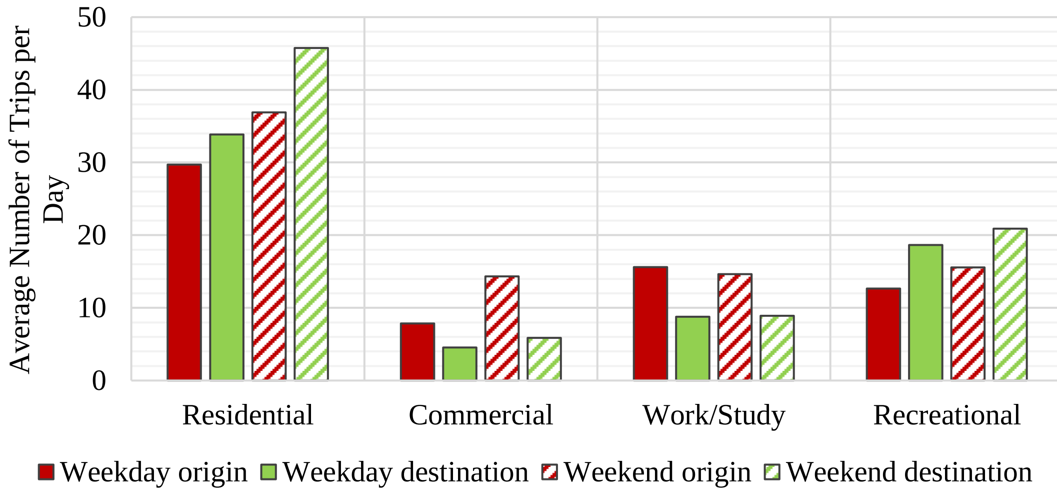


(c) Hourly variations.

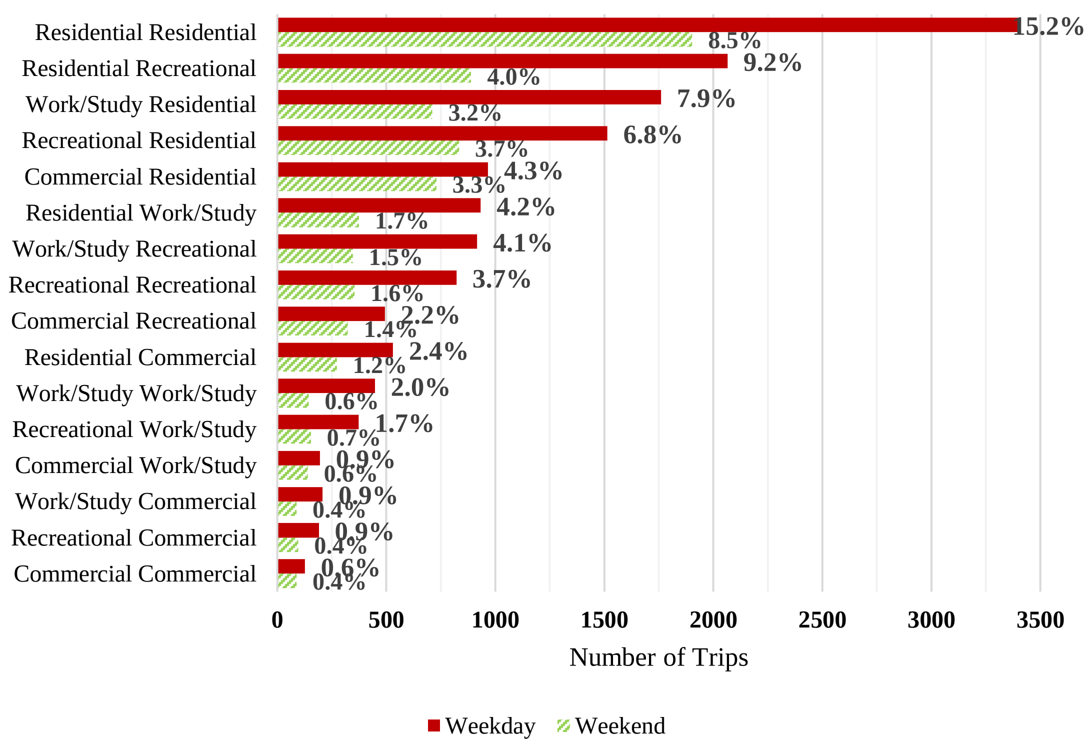
Figure 5.3: Daily, seasonal and hourly variations of taxi trips

The total travelled distance is around 321,240 kilometres. Around 20 % of taxi trips are shorter than 5 km; another 20 % have a length of 5 to 10 km; 35 % consist of trips from 10 to 20 km long; and the remaining 25 % are longer than 20 km. The mean, median and standard deviation of trip lengths are 14.3, 11.8, and 10.2 km, respectively. Although these values do not show any significant differences between trips made on weekdays compared to weekends, night trips and off-peak trips are slightly longer than day trips and peak hour trips, respectively. Another notable trend is that highway usage is considerably higher for night trips and off-peak trips, which is expected given lower congestion during these periods.

Figure 5.4(a) illustrates the average number of trips per day, classified based on the land use specification of the TAZ where the passenger was picked-up and dropped-off. Most trips originate from and/or have a destination in residential areas. It also highlights the high number of residential-end taxi trips for trips towards Laval. Trips are grouped based on OD land uses and the percentage of trips in each group is shown in Figure 5.4(b). Number of trips are represented by the length of the bar and labels represent the total share in percentages. A great majority of trips (around 46 %) were residential-based, while around 20 % start from work/study regions.



(a) Average number of trips per weekday/weekend per land use.



(b) Number and percentage of trips based on OD land uses.

Figure 5.4: Statistics on trips' land uses

5.5.2 Driver Characteristics

The dataset used in this study includes a subset of 1,746 drivers, comprising 204 night drivers and 498 day drivers, who only operate during nights and days, respectively. 68 drivers only operate during off peak hours, while 551 drivers work exclusively during peak hours. Around 38 % of peak hour trips take place during morning peak hours and the remaining 62 % occur during evening peak hours. An average of 13 trips and a median of seven trips per year (between the islands of Montreal and Laval) are recorded for each driver, and the maximum number is 286.

Taking a closer look at trips between same TAZ pairs reveals that certain drivers are more frequent among certain OD pairs, while others have a higher diversity of travelled TAZ pairs. More than 77 % of drivers make a single trip, and only around 5 % of them make four or more trips between a given TAZ pair. Among these more frequent drivers, 52 drivers make more than four trips between more than one pair of TAZ. The maximum number of trips between the same pair of TAZ by the same driver is found to be 40 trips.

5.5.3 Cluster Characteristics

As discussed in the previous section, a two-step clustering approach is adopted, in which PCA analysis precedes the clustering step. To assess the appropriateness of attributes' interrelationships and the usefulness of performing a PCA analysis, two measures have been evaluated: Bartlett's sphericity test (Bartlett, 1950; Bartlett, 1951), and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (Kaiser & Rice, 1974). The former measure, tests the hypothesis that the correlation matrix comes from an independent population of variables. The rejection of the independence hypothesis ($p < 0.05$) indicates that the data is appropriate for factor analysis (Hair, 2009). The latter measure, KMO, provides an index between 0 and 1 that measures the existing variance among variables. According Kaiser and Rice (1974), datasets with KMO values more than 0.5 are amenable to factor analysis. In this study, the null hypothesis of Bartlett's sphericity test has been rejected ($p < 0.05$), implying that the PCA can be efficiently performed on our dataset. In addition the KMO value is found to be 0.713 which according to Kaiser and Rice (1974) is 'middling', meaning that applying factor analysis may be beneficial.

The number of principal components considered in the clustering step, is defined based on four criteria. First, the Kaiser-Guttman criterion has been used (Cramer & Howitt, 2004; Guttman, 1954). This criterion argues that components with eigenvalues[3] greater than 1 should be retained. Second, the Karlis – Saporta – Spinaki criterion (Saporta, 2003) has been tested in order to select components that have eigenvalues significantly higher than 1, based on dataset characteristics such as the number of variables and observations. Third, the scree plot proposed by Cattell (1966) has been verified. In this plot, the eigenvalue of each component is plotted in descending order against the number of factors to visualize the amount of variance accounted for, by each component. Ideally, the last significant drop in eigenvalues divides the most important components from less important ones. Factors before the endpoint of this drop, known as the 'elbow' of the plot, are recommended to be retained (Cattell, 1966; Cramer & Howitt, 2004). Last, we used a parallel analysis method proposed by Horn (1965) in which eigenvalues are compared to those obtained from uncorrelated random normal variables, through a Monte Carlo simulation process. Factors with an eigenvalue bigger than those associated to random variables are considered to be significant. This technique is among the most accurate methods to determine the number of components (Glorfeld, 1995; Ledesma & Valero-Mora, 2007).

Based on these four criteria, the first two components have been extracted and used for the cluster analysis. These components have eigenvalues of 1.94 and 1.56, which are greater than 1 and 1.11, satisfying the Kaiser-Guttman and Karlis – Saporta – Spinaki criteria, respectively. The substantial difference between the eigenvalues of the second and third components (1.56 and 0.96, respectively), the examination of the scree plot and the parallel analysis also suggest the selection of the first two components. These latent components explain more than 60 % of the variation between attributes implying that they carry most of the attitudinal information (Fu & Juan, 2016).

A major challenge in clustering analysis is the selection of an appropriate number of clusters. Although visualization may be an effective way to verify results, it is highly difficult to visualize data with more than three dimensions. In this study, a series of two to eight clusters have been experimented and the optimal number of clusters has been defined by the maximum value of Between-group Sum of Squares (BSS ratio) and GAP-statistic (Everitt, Stahl, Leese, & Landau, 2011), as well as the behavioural interpretation associated with each cluster.Furthermore, the dendrogram representing the HAC process, has been inspected to verify the plausibility of results.

---

[3] Here the eigenvalue of a component refers to the amount of variance explained by a component.

Considering the two selected principal components, a set of four clusters has been found to provide the best results. Results showed significant variations of drivers' behaviours towards shorter versus longer routes, and routes taken during day versus night. Interestingly, attributes such as the exact hour of the trip, variation of trip days, seasonal variations, and different OD land uses showed no significant impacts on drivers' route choice behaviours. To measure results' stability, first, the dataset was randomly divided in two halves and the clustering process was separately performed on each half for the same parameter settings; and second, observations were randomly permuted in our dataset. Since results were not significantly different, it was concluded that the four-cluster solution has a high degree of stability and reliability (Mooi & Sarstedt, 2010).

Some descriptive statistics of attributes, based on which the final clustering has been described, are illustrated in Table 5.2. Correlations between these attributes are presented in Table 5.3. Clusters are presented in Table 5.4 and characterized through their significant attributes. To assess attributes' level of significance, the Test Value (TV) criterion is evaluated.

Table 5.2: Descriptive analysis of significant factors in the clustering step

Attribute	**Mean (SD)**	**Median**	**SD**[a]	**Min.**	**Max.**
Day	57.3	60.0	36.5	0.0	100.0
Weekday	64.4	66.7	27.9	0.0	100.0
Peak	24.3	20.0	25.5	0.0	100.0
Length (km)	18.5	18.3	7.3	1.2	65.6
Travel time (min)	21.4	20.9	7.3	1.2	62.4
% Highway	54.4	56.7	20.5	0.0	100.0
Path-Size	1.0	1.0	0.1	0.4	1.0
# Links/km	5.7	5.6	1.3	2.5	12.3
# Turns/km	0.6	0.6	0.3	0.0	2.4
% Shortest path	43.9	43.1	18.2	0.0	99.4
# TAZ	12.7	8.0	15.3	1.0	154.0
TAZ_4	0.2	0.0	1.2	0	23
TAZ_max	1.6	1	2.2	1	40

[a] Standard Deviation

Table 5.3: Correlation matrix of significant factors in the clustering step

Attribute		1	2	3	4	5	6	7	8	9	10	11	12	13
1	Day	-												
2	Weekday	0.31	-											
3	Peak	0.61	0.17	-										
4	Length	-0.11	-0.03	-0.07	-									
5	Travel time	0.08	0.06	0.08	0.69	-								
6	% Highway	-0.08	-0.01	-0.07	0.65	0.19	-							
7	Path-Size	-0.06	-0.09	-0.05	0.20	0.25	0.13	-						
8	# Links/km	0.02	-0.05	0.05	-0.58	-0.14	-0.92	-0.08	-					
9	# Turns/km	0.18	0.07	0.15	-0.35	-0.12	-0.61	-0.14	0.66	-				
10	% SP [a]	0.15	0.08	0.08	-0.39	-0.38	-0.17	-0.12	0.10	-0.05	-			
11	# TAZ	-0.03	0.06	-0.01	-0.35	-0.37	-0.20	-0.59	0.15	0.22	0.08	-		
12	TAZ_4	0.00	0.02	0.00	-0.20	-0.20	-0.17	-0.61	0.14	0.11	0.08	0.63	-	
13	TAZ_max	0.03	0.05	0.02	-0.24	-0.27	-0.18	-0.79	0.14	0.15	0.10	0.69	0.84	-

[a] Shortest Path

Table 5.4: Clusters specifications

	Short trip night drivers		Long trip night drivers		Short trip day drivers		Long trip day drivers		All
	TV [a]	Mean (SD[b])	TV	Mean (SD)	TV	Mean (SD)	TV	Mean (SD)	Mean (SD)
Day	-22.0	22.5 (20.6)	-15.8	32.2 (27.5)	10.9	75.8 (27.6)	24.6	88.4 (16.4)	57.3 (36.5)
Weekday	-12.8	48.9 (26.4)	-7.6	55.2 (26.9)	3.5	69.0 (25.2)	15.3	79.2 (22.5)	64.4 (27.9)
Peak	-15.5	7.3 (10.0)	-13.9	9.0 (11.1)	6.5	32.0 (24.7)	20.9	42.7 (26.3)	24.3 (25.5)
Length (km)	-6.8	16.4 (3.9)	19.7	24.8 (5.3)	-24.6	10.2 (3.3)	9.9	20.1 (6.6)	18.5 (7.3)
Travel time	-4.0	20.2 (5.9)	6.1	23.3 (5.3)	-16.4	15.9 (6.1)	12.5	24.5 (7.8)	21.4 (7.3)
% Highway	-7.2	48.0 (13.8)	19.0	71.3 (9.9)	-25.6	30.1 (16.2)	11.8	62.7 (15.1)	54.4 (20.5)
Path-Size	-3.9	1.0 (0.04)	6.4	1.0 (0.03)	-13.9	0.9 (0.1)	2.8	1.0 (0.1)	1.0 (0.1)
# Links/km	8.2	6.2 (1.0)	-16.6	4.7 (0.7)	22.6	6.1 (1.2)	-12.1	5.1 (1.0)	5.7 (1.3)
# Turns/km	10.4	0.8 (0.2)	-12.0	0.5 (0.2)	17.8	0.9 (0.3)	-4.3	0.6 (0.2)	0.7 (0.3)
% SP [c]	3.2	48 (16.8)	-7.9	37.7 (17.3)	10.5	52.8 (17.7)	-2.9	38.4 (17.9)	44.0 (18.2)
# TAZ	-1.9	11.5 (9.8)	-5.4	9.1 (6.9)	15.3	23.5 (26.0)	-6.8	9.1 (9.0)	12.7 (15.3)
TAZ_4	-3.0	0.1 (0.2)	-3.4	0.1 (0.2)	10.3	0.8 (2.5)	-3.4	0.1 (0.3)	0.2 (1.2)
TAZ_max	-3.6	1.3 (0.7)	-4.9	1.2 (0.5)	13.5	3.0 (4.4)	-4.1	1.3 (0.9)	1.6 (2.2)
# Obs. (%)	408 (23.4 %)		405 (23.2 %)		368 (21.1 %)		565 (32.3 %)		1746 (100 %)

[a] Test Value [b] Standard Deviation [c] Shortest Path

First, a Temporal Stratification is observed, separating drivers working during nights and days. Then, an Operational Stratification further divides each group into two separate categories, forming four behavioural clusters. These clusters are labelled: Short trips night drivers, Long trips night drivers, Short trips day drivers, and Long trips day drivers, and represent 23.4 %, 23.2 %, 21.1 %, and 32.3 % of our sample population, respectively. By looking at the Temporal Stratification of drivers, it is observed that contrary to day drivers who operate more on weekdays and peak hours, night drivers operate more on weekends and off-peak hours. Moreover, highway usage is considerably higher for night trips. In other words, night drivers are more likely to use highways for both their long and short trips, compared to day drivers. This may be due to higher congestion levels on highways, during the day. Trip lengths in night clusters have been found to be greater than day clusters, both for short and long trips. However, despite longer trip lengths, travel times are slightly shorter for night trips, due to higher congestion during these periods. Furthermore, day trips are more similar to their respective distance based shortest path alternatives than night trips. This may be due to the fact that, given a lower congestion level at night, drivers might take an alternative route that is faster or more familiar.

Moreover, the operational stratification of observed route choices results in two different types of driving behaviour. Drivers who most frequently take shorter trips usually do not use highways and prefer to take local routes with higher number of intersections and turns, which accentuate their familiarity with the road network and their awareness of traffic conditions. These drivers tend to choose similar routes between same OD pairs, which have higher proportions of identical segments with the distance-based shortest path. On the other hand, longer trip drivers showed higher propensity towards using highways to avoid intersections and turns and to benefit from higher speed limits, which is intuitive. These drivers revealed more willingness in taking more diverse routes between same OD pairs, which are less similar to the distance-based shortest path than routes taken in shorter trips. This might be partly due to the higher number of feasible alternatives between distanced OD pairs compared to closer ones.

Taking a closer look at the dataset demonstrates that regular drivers, i.e. those that have higher values of TAZ_4 and TAZ_max (which implies that they travel more between certain OD pairs), have a higher tendency towards short trips during the day. Since these drivers have a positive Test Value for the total number of unique TAZ pairs travelled (#TAZ), it can also be inferred that they have probably a better knowledge of the network. In order to visualize these clusters, the Z-score (cluster mean minus population mean, divided by population standard deviation), has been calculated (see Figure 5.5). This value indicates the relative position of a score to the population mean. Positive and negative values indicate scores above and below the mean, respectively.

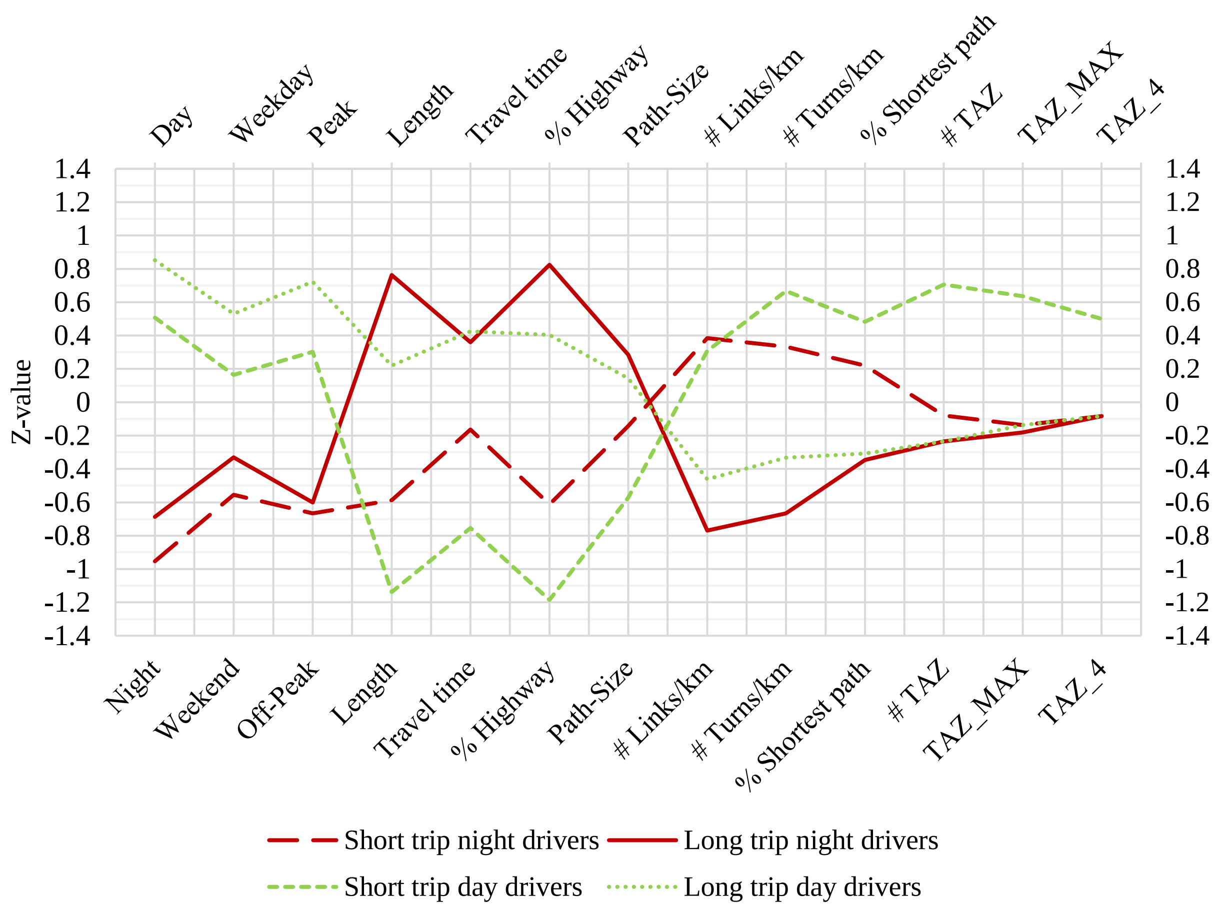


Figure 5.5: Profile plot of the four clusters

5.6 Conclusions

Although previous studies have shown that different categories of road users are observable, there is a lack of a representative classification of taxi drivers' operating strategies based on their actual route choices over a long duration of time. New possibilities have opened up due to the availability of very large and rich datasets from taxi companies that maintain detailed GPS trajectories of their fleets for a long duration of time.

The main objective of this research is to improve the understanding of taxi drivers' operating strategies, by classifying them based on their observed route choices, using longitudinal GPS datasets. In this effort, a dataset comprising more than 22,000 trips, made by 1,746 taxi drivers over a period of one year, for trips originating in Montreal with a destination in Laval, has been studied. It is worth mentioning that the destination choice is not the focus of this research and this specific regional context has been chosen to ensure a wide range of travel distances.

We first presented important statistical properties of taxi trips. Since various degrees of correlation were observed among attributes, a PCA analysis has been performed to improve the efficiency of the clustering algorithm and to reduce the probability of the algorithm getting stuck in a local optima (Poucin et al., 2016). Then, a HAC algorithm has been performed to classify drivers' behaviours.

Due to the significant behavioural variations found in trips made during days and nights, and between short trips and long trips, these four clusters were labelled *"Short trips night drivers"*, *"Long trips night drivers"*, *"Short trips day drivers"*, and *"Long trips day drivers"*. Intuitively, drivers prefer highways for longer trips and local routes with higher number of turns and intersections for shorter trips. In addition, results demonstrated that drivers tend to take more similar routes between same OD pairs, where the travelled distance is short. It is also perceived that routes taken between closer OD pairs are more similar to the distance-based shortest path compare to those taken between distanced OD pairs.

Although it is not possible to encompass all variations of operating strategies based on route choice behaviours and GPS traces alone, due to the lack of some other explanatory variables, such as demographics and preferences, this study shed some light on the variation of taxi drivers' operating strategies and route choice behaviours and the possibility of classifying them using clustering algorithms.

Understanding the behavioural heterogeneity in drivers' route choices is very important for city and transportation planners. Fu and Juan (2016) concluded that the stratification of a studied population according to their preferences would improve the effectiveness of transport measures and the efficiency of policy implication, since it better captures the heterogeneity among different segments of the population. The recent rise of ride-hailing services has pushed the cities and transportation planners to develop new laws and policies for urban mobility. Montreal, like many large cities, is experiencing the aggressive market penetration of transportation network companies like Uber, very strong resistance from taxi companies, and the rise of new local taxi services. The decision makers are faced with the challenge of developing new policies that are better for the residents as users and more sustainable for a free-market where conventional taxis and ride-hailing services can co-exist and grow. This study can contribute in proposing fair policies for the incumbent of taxi industry and other recent transportation companies, in order to ensure and encourage a shared economy.

Taxi drivers are considered as a well-informed group of drivers who have acquired higher knowledge of the road network and its travel time variations. From a policy and planning perspective, an improved understanding of taxi drivers' operating strategies and route choice behaviours is important, since taxi, as an important mode of transportation in big cities, provides further insights on human mobility patterns, traffic dynamics, and urban structures. Moreover, taxis contribute to the road congestion by occupying a considerable share of the limited road capacity both when they are vacant and in search for a passenger, and when they are occupied (Yang, Ye, Tang, & Wong, 2005). An improved understanding of taxis operating strategies and their route choice behaviours can help authorities and planning agencies to propose new policies or impose new regulations and restrictions to alleviate congestion in big cities.

Findings of this research pave the route for several future research directions. A major limitation of this work is the lack of personal information such as demographic and socio-economic variables. Similar longitudinal datasets including those variables can be used to enrich the findings with more personal information. A future direction would include regular car drivers and the validation of the four behavioural clusters proposed in this study. Since the experience level of regular car drivers varies widely from taxi drivers, different driving patterns may be expected. Another interesting elaboration of this study could be the incorporation of these clusters into route choice models in order to examine the enhancement of their estimation and prediction abilities.

Acknowledgments

The authors would like to thank collaborators from Taxi Diamond who provided access to the data for research purposes. They are grateful to Jean-Simon Bourdeau, as well as Annick Lacombe for their help in extracting the raw data needed for this study. They also acknowledge the Mobilité Chair and its partners for the financial support of this project.

References

Alizadeh, H., Farooq, B., Morency, C., Saunier, N.: On the role of bridges as anchor points in route choice modeling. Transportation, 1-26 (2017a). doi:10.1007/s11116-017-9761-7

Alizadeh, H.F., Bilal , Morency, C., Saunier, N.: Classifying Behavioral Dynamics of Taxi Drivers Route Choices Using Longitudinal GPS Data. In: Transportation Research Board 96th Annual Meeting, Washington, D.C. 2017b

Bartlett, M.S.: Tests of significance in factor analysis. British Journal of statistical psychology 3(2), 77-85 (1950)

Bartlett, M.S.: A further note on tests of significance in factor analysis. British Journal of Statistical Psychology 4(1), 1-2 (1951)

Ben-Akiva, M., Bierlaire, M.: Discrete choice methods and their applications to short term travel decisions. In: Handbook of transportation science. pp. 5-33. Springer, (1999)

Ben-Akiva, M., De Palma, A., Isam, K.: Dynamic network models and driver information systems. Transportation Research Part A: General 25(5), 251-266 (1991)

Bovy, P.H., Stern, E.: Route choice: Wayfinding in transportation networks. In. Kluwer Academic Publishers, Dordrecht, (1990)

Cattell, R.B.: The scree test for the number of factors. Multivariate behavioral research 1(2), 245-276 (1966)

Communauté métropolitaine de montréal, C.: An Attractive, Competitive and Sustainable Greater Montreal. In. Library and Archives Canada Montréal, Québec, (2012)

Cramer, D., Howitt, D.L.: The Sage dictionary of statistics: a practical resource for students in the social sciences. Sage, (2004)

Damant-Sirois, G., Grimsrud, M., El-Geneidy, A.M.: What's your type: a multidimensional cyclist typology. Transportation 41(6), 1153-1169 (2014). doi:10.1007/s11116-014-9523-8

Dill, J., McNeil, N.: Four types of cyclists? Examining a topology to better understand bicycling behavior and potential. In: 92nd Annual Meeting of the Transportation Research Board, Washington, DC 2013

Eagle, N., Pentland, A.S.: Eigenbehaviors: identifying structure in routine. Behavioral Ecology and Sociobiology 63(7), 1057-1066 (2009). doi:10.1007/s00265-009-0739-0

Everitt, B.S., Stahl, D., Leese, M., Landau, S.: Cluster analysis. (2011)

Fu, X., Juan, Z.: Empirical analysis and comparisons about time-allocation patterns across segments based on mode-specific preferences. Transportation 43(1), 37-51 (2016). doi:10.1007/s11116-014-9561-2

Geller, R.: Four types of cyclists. PortlandOnline, (2009)

Glorfeld, L.W.: An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. Educational and psychological measurement 55(3), 377-393 (1995)

Guttman, L.: Some necessary conditions for common-factor analysis. Psychometrika 19(2), 149-161 (1954). doi:10.1007/bf02289162

Hair, J.F.: Multivariate data analysis. (2009)

Higgs, B., Abbas, M.: A two-step segmentation algorithm for behavioral clustering of naturalistic driving styles. In: 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013) 2013, pp. 857-862. IEEE

Horn, J.L.: A rationale and test for the number of factors in factor analysis. Psychometrika 30(2), 179-185 (1965). doi:10.1007/bf02289447

Jensen, M.: Passion and heart in transport—a sociological analysis on transport behaviour. Transport Policy 6(1), 19-33 (1999)

Kaiser, H.F., Rice, J.: Little Jiffy, Mark IV. Educational and psychological measurement (1974)

Kang, C., Qin, K.: Understanding operation behaviors of taxicabs in cities by matrix factorization. Computers, Environment and Urban Systems 60, 79-88 (2016). doi:http://dx.doi.org/10.1016/j.compenvurbsys.2016.08.002

Kroesen, M., Handy, S.: The relation between bicycle commuting and non-work cycling: results from a mobility panel. Transportation 41(3), 507-527 (2014). doi:10.1007/s11116-013-9491-4

Lacombe, A., Morency, C.: Modeling Taxi Trip Generation Using GPS Data: The Montreal Case. In: Transportation Research Board 95th Annual Meeting 2016, vol. 16-4345

Lebart, L., PIRON, A., Lebart, M., Morineau, A., Piron, M.: Statistique exploratoire multidimensionnelle. Dunod, (2000)

Ledesma, R.D., Valero-Mora, P.: Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. Practical assessment, research & evaluation 12(2), 1-11 (2007)

Liu, L., Andris, C., Ratti, C.: Uncovering cabdrivers' behavior patterns from their digital traces. Computers, Environment and Urban Systems 34(6), 541-548 (2010)

Lucidi, F., Giannini, A.M., Sgalla, R., Mallia, L., Devoto, A., Reichmann, S.: Young novice driver subtypes: relationship to driving violations, errors and lapses. Accident Analysis & Prevention 42(6), 1689-1696 (2010)

Manley, E., Addison, J., Cheng, T.: Shortest path or anchor-based route choice: a large-scale empirical analysis of minicab routing in London. Journal of Transport Geography 43, 123-139 (2015)

Marengo, D., Settanni, M., Vidotto, G.: Drivers' subtypes in a sample of Italian adolescents: relationship between personality measures and driving behaviors. Transportation research part F: traffic psychology and behaviour 15(5), 480-490 (2012)

Mooi, E., Sarstedt, M.: Cluster analysis. Springer, (2010)

Morency, C., Trepanier, M., Agard, B.: Typology of carsharing members. In: Transportation Research Board 90th Annual Meeting 2011, vol. 11-1236

Okamoto, K., Utsumi, A., Yamazoe, H., Miyashita, T., Abe, S., Takahashi, K., Hagita, N.: Classification of Pedestrian Behavior in a Shopping Mall based on LRF and Camera Observations. In: MVA 2011, pp. 1-5

Pagès, J.: Analyse factorielle de données mixtes. Revue de statistique appliquée 52(4), 93-111 (2004)

Parkany, E., Du, J., Aultman-Hall, L., Gallagher, R.: Modeling stated and revealed route choice: consideration of consistency, diversion, and attitudinal variables. Transportation Research Record: Journal of the Transportation Research Board(1985), 29-39 (2006)

Pele, N., Morency, C.: When, Where and How Taxis Are Used in Montreal. In: Transportation Research Board 93rd Annual Meeting 2014, vol. 14-3943

Poucin, G., Farooq, B., Patterson, Z.: Pedestrian Activity Pattern Mining in WiFi-Network Connection Data. In: Transportation Research Board 95th Annual Meeting 2016, vol. 16-5846

Rakotomalala, R.: TANAGRA: un logiciel gratuit pour l'enseignement et la recherche. In: EGC 2005, pp. 697-702

Ramaekers, K., Reumers, S., Wets, G., Cools, M.: Modelling route choice decisions of car travellers using combined GPS and diary data. Networks and Spatial Economics 13(3), 351-372 (2013)

Reinoso, G., Farooq, B.: URBAN PULSE ANALYSIS USING BIG DATA. Paper presented at the 50th Annual Conference of Canadian Transportation Research Forum,

Saporta, G.: A control chart approach to select eigenvalues in principal component and correspondence analysis. 54th Session of the International Statistical Institute-Berlin (2003)

Saunier, N., Mourji, N., Agard, B.: Mining microscopic data of vehicle conflicts and collisions to investigate collision factors. Transportation Research Record: Journal of the Transportation Research Board(2237), 41-50 (2011)

Shaaban, K., Kim, I.: Assessment of the taxi service in Doha. Transportation Research Part A: Policy and Practice 88, 223-235 (2016). doi:http://dx.doi.org/10.1016/j.tra.2016.04.011

Tawfik, A.M., Rakha, H.A., Miller, S.D.: Driver route choice behavior: Experiences, perceptions, and choices. In: Intelligent Vehicles Symposium (IV), 2010 IEEE, 21-24 June 2010 2010a, pp. 1195-1200

Tawfik, A.M., Rakha, H.A., Miller, S.D.: An experimental exploration of route choice: identifying drivers choices and choice patterns, and capturing network evolution. In: Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on 2010b, pp. 1005-1012. IEEE

Tawfik, A.M., Szarka, J., House, L., Rakha, H.A.: Disaggregate route choice models based on driver learning patterns and network experience. In: 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), 5-7 Oct. 2011 2011, pp. 445-450

Ulleberg, P.: Personality subtypes of young drivers. Relationship to risk-taking preferences, accident involvement, and response to a traffic safety campaign. Transportation Research Part F: Traffic Psychology and Behaviour 4(4), 279-297 (2001). doi:http://dx.doi.org/10.1016/S1369-8478(01)00029-8

Yang, H., Ye, M., Tang, W.H., Wong, S.C.: Regulating taxi services in the presence of congestion externality. Transportation Research Part A: Policy and Practice 39(1), 17-40 (2005)

CHAPTER 6 ARTICLE 3: AN ONLINE SURVEY TO ENHANCE THE UNDERSTANDING OF CAR DRIVERS ROUTE CHOICES

Hamzeh Alizadeh[1*], Pierre-Léo Bourbonnais[1], Catherine Morency[1], Bilal Farooq[2], Nicolas Saunier[1]

Accepted for publication in *Transportation Research Procedia: 11th International Conference on Transport Survey Methods*

Submission Date: 3 / October / 2017

[1] Department of Civil, Geological and Mining Engineering, Polytechnique Montréal, C.P. 6079, succ. Centre-Ville Montréal, Canada

[2] Department of Civil Engineering, Ryerson University, 350 Victoria Street, Toronto, Canada

[*] Corresponding author

Abstract

The increasing application of advanced behavioural choice models, reflecting the stochasticity of individuals' preferences and the complex nature of human decision-making process, requires enhanced data collection methods to obtain detailed data without significantly increasing the respondent burden. In this study, we present the development and deployment of a general data collection framework adapted for behavioural route choice studies. The main objectives of the proposed framework are to observe drivers' route choices, and to identify important factors, including observable attributes and latent behavioural traits, affecting those decisions. More specifically, the survey has been designed to reveal drivers' consideration set of route alternatives from which they pick their final choices. The detailed analysis of survey's response behaviour will help improve the framework to gather travel data even more efficiently.

Keywords: Revealed Preference Survey, Route Choice Modelling, Consideration Sets, Survey Response Behaviour

6.1 Introduction

Demand in a road network, i.e. the traffic flow, is the direct result of individuals' travel and route choice decisions. These decisions are captured through route choice models, which are considered as an important part of traffic assignment procedures, and play a key role in transportation planning to forecast the traffic flow, design new transportation infrastructures, and investigate new policies. It is therefore of utmost importance to understand drivers' route choice behaviours and factors affecting them.

Route choice modelling investigates the process of route selection by an individual, making a trip between predefined origin-destination (OD) pairs, and is probably one of the most complex and challenging problems in traffic assignment. The complexity of this process is partly due to factors such as the sophisticated nature of human behaviour, the ambiguity of the decision-making process, and the stochasticity of individuals' preferences.

Previous route choice studies have mostly focused on the effects of *observable* factors on drivers' route choice decisions (Alizadeh, Farooq, et al., 2017b; Dalumpines & Scott, 2017b). Instances of these factors include route features (i.e. travel distance, number of turns, etc.) and drivers' characteristics (such as age, gender, etc.), which are tangible and can be directly observed. However, numerous studies have shown that *latent* factors such as attitudinal traits, perceptions and lifestyle preferences play a major role in the decision-making process (Ben-Akiva et al., 2002; Prato et al., 2012; Walker, 2001). Attitudes can be defined as the tendency of having a favorable or unfavorable evaluation towards something (Bradley, 2013). Perception, however, can be thought of as a subjective subconscious process of interpretation of events, experiences, and stimuli, which can be influenced and shaped by previous occurrences (Bradley, 2013). For instance, in addition to the observable factors, drivers' decisions might also be influenced by factors such as safety concerns, driving habits and spatial abilities (Gärling et al., 1998; Kamargianni, Dubey, Polydoropoulou, & Bhat, 2015; McFadden, 1999; Muñoz, Monzon, & López, 2016; Sarkar & Mallikarjuna, 2017).

The complexity of route choice modelling is also attributed to the high density of the road network, the large number of possible alternatives between OD pairs, and the correlations among these alternatives. Since it is not computationally feasible to enumerate all the possible routes connecting a given OD pair (i.e. the *universal choice set*) in a real world road network, nor behaviourally accurate to assume that drivers are aware of all of them, a two stage choice modelling process is usually adopted (Manski, 1977). In the first stage of this process, a subset of the universal choice set is selected to form the collection of feasible travel alternatives considered by the driver (i.e. the *considered choice set*). Then, in the second stage, drivers make their final route choices from the considered set of route alternatives.

However, defining a proper consideration set is a serious challenge in route choice modelling. The consideration set of route alternatives is usually latent to the analyst and alternatives are therefore usually generated using variations of the shortest path algorithm (Hoogendoorn-Lanser et al., 2005; Prato et al., 2012). The generated set should include alternatives that are attractive to the driver in a real world choice situation (Hess & Daly, 2010), and the misspecification of the size and composition of the considered choice set greatly affect model's estimates and may lead to fallacious predicted demand levels (Bliemer & Bovy, 2008; Geda, 2014; Peters et al., 1995; Prato & Bekhor, 2006, 2007b; Schuessler & Axhausen, 2009; Swait & Ben-Akiva, 1987a).

Considering the abovementioned challenges of route choice modelling, and in order to improve the estimation and prediction of drivers' route choice decisions, it is imperative to first, observe drivers' revealed preferences in real route choice situations, second, to identify behavioural and attitudinal factors as additional sources of heterogeneity affecting their decisions, and finally, to get a better grasp of the formation process, size and composition of drivers' considered sets of route alternatives. Accordingly, this study proposes a framework of data collection for route choice studies, with the objective of satisfying the aforementioned requirements.

The remainder of the paper is structured as follows. First, we briefly review some of the previous route choice studies and their data collection methods, and in that context further clarify the contributions of the presented data collection framework. Next, we present the proposed survey framework and its implementation. Survey participants, their response behaviours, completion rates and dropouts are discussed in the next section. In the end, we highlight the possible implications and applications of this survey framework, underscore its limitations, and suggest further research directions.

6.2 Previous Studies

Table 6.1 summarizes the data collection methods adopted in some of the previous route choice studies and enumerates the attributes that have been found to significantly affect drivers' decisions. Also, it puts into perspective the specifications and characteristics of the proposed framework. Although the table does not encompass all the previous route choice studies, this list has been selected to provide a wide spectrum of research on that matter. Studies have been compared using four criteria, namely *Medium*, *Method*, *Collected data*, and *Significant attributes*.

Medium refers to the type of interface that has been used to collect the data. Different types of media have been used during the past few decades. Telephone, mail, face-to-face, and web surveys are among the typical methods that have been extensively used to collect accurate choice data. In order to reduce the respondent burden, computer-assisted self / telephone interviewing have been adopted (Papinski, 2011; Srinivasan & Dhakar, 2013). In Table 6.1, four types of interfaces have been identified for route choice data collection. Household Travel Survey (HTS) data, as a traditional source of data, has been used in few studies. Such diaries are not very effective in collecting detailed, long-term or large scale route choice data, due to the excessive respondent burden of declaring the exact routes (Chen, 2013b). Therefore, the detailed trajectory is usually not available in HTS data, and a shortest path algorithm (based on some generalized cost function) is used to simulate the chosen route. Paper-based (PB), computer-based (CB), and web-based (WB) route choice surveys are among other types of data collection media adopted in route choice studies.

Method indicates the methodology to observe and quantify respondents' preferred choice. Route choice surveys are either Revealed Preference (RP), in which respondents reveal their actual choices in real route choice situations, or Stated Preference (SP), in which respondents are asked to choose between several hypothetical route alternatives based on some provided details on each choice. In recent years, the prevalent use of GPS technology has provided researchers with an abundance of high-resolution geospatial data. An important advantage of GPS data collection over other methods is that it can record travel information for several days without any additional respondent burden. However, working with GPS data brings several complexities including the large size of the dataset, the absence of data due to signal loss, the challenges in constructing a representative sample, and the technological issues such as battery life and record accuracy amongst others. Furthermore, studies based on GPS data often lack personal information on the decision maker, his attitudes, experiences, and preferences. Even though GPS traces can be considered as RP data, we considered them as a separate method of data collection in Table 6.1, due to their completely different nature and data processing requirements, and their prevalence of use in route choice studies.

Regardless of the method used, the main objective of all the above-mentioned data collection methods is to record the observed choices (Obs.). The other types of collected data depends to a large extent on the objectives of the survey. For instance, to make an in-depth analysis of the effect of behavioural traits (Behvr.) on the final choice, attitudinal questions and psychometric indicators should be the focus of the survey, while to analyze respondents' perception bias towards a particular choice, questions regarding the perceived values of different attributes (Percp.) is of prior importance. It is also a common procedure to ask respondents to reveal the most important factors affecting their choices (Fact.).

The observation of the considered choice set (CCS) of route alternatives, from which drivers make their final choices, is not as straightforward, and hence, is not very common in practice. In SP surveys, participants make their choices from a series of hypothetical alternatives provided by the analyst, while in GPS surveys, the consideration set of route alternatives mostly remains unidentified. Moreover, the specification of the considered choice set is usually ignored in RP surveys to reduce the response time as well as the respondents' burden.

Finally, factors that have been found to significantly affect drivers' route choice process is compared in the last column of Table 6.1 (i.e., Significant attributes). The variety of factors that have been found to significantly affect drivers' route choice behaviour further illustrates the importance that survey design should be in line with the objectives of the survey and the expected results.

Table 6.1: Comparison of selected route choice studies and their data collection methods

Study	Medium				Method			Collected data						Significant Attributes *
	HTS[a]	PB[b]	CB[c]	WB[d]	RP[e]	SP[f]	GPS	Obs[g]	CCS[h]	Demo[i]	Fact[j]	Behvr[k]	Percp[l]	
Iida et al. (1994)	☐	☐	☒	☐	☐	☒	☐	☒	☐	☐	☐	☐	☐	28, 30
Abdel-Aty and Jovanis (1997)	☐	☒	☒	☐	☒	☒	☐	☒	☐	☐	☒	☐	☒	1, 2, 3, 4, 5, 6
Peeta et al. (2000)	☐	☒	☐	☐	☐	☒	☐	☒	☐	☒	☐	☒	☐	9, 12, 28
Cascetta et al. (2002)	☐	☒	☐	☐	☒	☐	☐	☒	☒	☒	☐	☐	☐	1, 12, 17
Parkany et al. (2006b)	☐	☒	☐	☐	☐	☐	☒	☒	☐	☒	☒	☒	☐	3, 9, 10
Cools et al. (2009)	☒	☐	☐	☐	☒	☐	☐	☒	☐	☒	☐	☐	☐	1, 9, 10, 23
Papinski et al. (2009)	☐	☒	☒	☒	☐	☐	☒	☒	☐	☒	☒	☐	☐	1, 3, 4, 6, 21, 22
Ben-Elia and Shiftan (2010)	☐	☐	☒	☐	☐	☒	☐	☒	☐	☒	☐	☐	☐	14, 15, 25, 29
Tawfik et al. (2010b)	☐	☒	☐	☐	☒	☐	☐	☒	☐	☒	☐	☒	☒	1, 6, 9, 15, 24, 25
Schlaich (2010)	☐	☐	☐	☐	☐	☐	☒	☒	☐	☐	☐	☐	☐	3, 28
Prato et al. (2012)	☐	☐	☐	☒	☒	☐	☐	☒	☒	☒	☐	☒	☐	1, 5, 6, 7, 9, 13, 27
Kaplan and Prato (2012)	☐	☐	☐	☒	☒	☐	☐	☒	☐	☒	☐	☒	☐	1, 5, 6, 7, 9, 13, 27
Gan and Chen (2013)	☐	☐	☐	☐	☐	☒	☐	☒	☐	☒	☐	☐	☐	14, 15, 16
Jou and Yeh (2013)	☐	☐	☐	☐	☐	☒	☐	☒	☐	☒	☐	☐	☐	1, 5, 6, 8, 10, 12, 18, 19
Tawfik and Rakha (2013)	☐	☒	☐	☐	☒	☐	☐	☒	☐	☐	☐	☒	☒	1, 6, 9, 15, 24, 26, 27
Ramaekers et al. (2013)	☐	☒	☐	☐	☐	☐	☒	☒	☐	☒	☒	☐	☐	8, 9, 10, 11
Koller-Matschke et al. (2013)	☐	☐	☐	☒	☒	☒	☒	☒	☐	☐	☐	☐	☐	1, 18, 28, 29
Habib et al. (2013)	☒	☐	☐	☐	☒	☐	☐	☒	☐	☒	☐	☐	☐	1, 8, 10, 20
Vacca and Meloni (2014)	☐	☐	☐	☐	☐	☐	☒	☒	☐	☒	☐	☐	☐	1, 5, 9, 12, 15, 22, 24
Hess et al. (2015)	☐	☐	☐	☐	☐	☐	☒	☒	☐	☐	☐	☐	☐	1, 6, 11, 19
Manley, Addison, et al. (2015b)	☐	☐	☐	☐	☐	☐	☒	☒	☐	☐	☐	☐	☐	20
Lai and Bierlaire (2015)	☐	☐	☐	☐	☐	☐	☒	☒	☐	☐	☐	☐	☐	5, 6, 22
Dalumpines and Scott (2017b)	☐	☐	☐	☐	☐	☐	☒	☒	☐	☒	☐	☐	☐	1, 5, 6, 7, 11
Alizadeh, Farooq, et al. (2017b)	☐	☐	☐	☐	☐	☐	☒	☒	☐	☐	☐	☐	☐	4, 5, 6, 20
Proposed Framework	☐	☐	☐	☒	☒	☐	☐	☒	☒	☒	☒	☒	☒	-

***Factors affecting route choice**

1- Travel time	7- Number of turns	13- Delay	19- Toll rate / Cost	25- Learning process
2- Travel time reliability	8- Time of day	14- Network familiarity	20- Anchor points	26- Drivers' categories
3- Traffic conditions (level of service)	9- Socio-demographic	15- Driving experience	21- Stop signs	27- Personality traits
4- Number of segments	10- Trip purpose	16- Education	22- Traffic lights	28- Availability of information
5- Percentage of highway	11- Road type	17- Topology	23- Holidays	29- Type of information
6- Travel distance	12- Socio-economic	18- Choice inertia	24- Travel speed	30- Quality of information

[a] Household Travel Survey [b] Paper Based [c] Computer Based [d] Web based [e] Revealed Preference [f] Stated Preference [g] Observed Choices
[h] Considered Choice Set [i] Demographics [j] Revealed Factors [k] Behavioural traits [l] Perception

6.3 Survey Framework

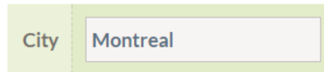
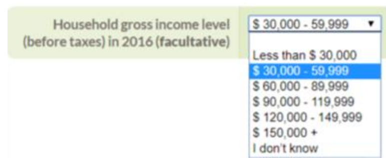
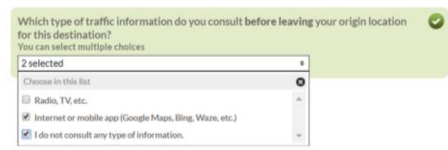
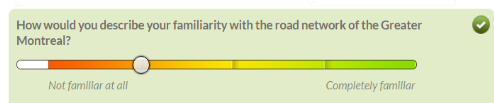
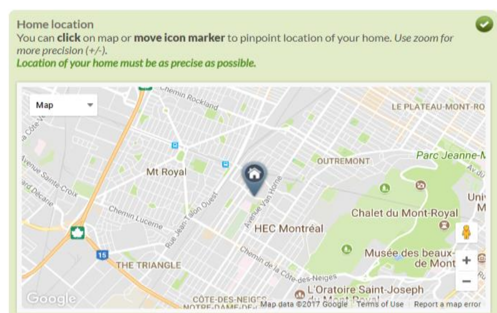
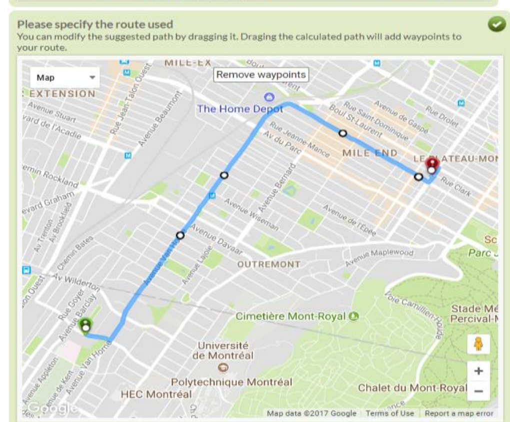
In this section, we present the development and implementation of the proposed revealed preference web-based survey, designed to observe drivers' revealed route choices towards their most frequently visited destinations, and identify behavioural and attitudinal factors affecting them. We also intend to observe drivers' consideration sets of route alternatives and to characterize them based on drivers' perceptions.

Drivers residing and driving in the Greater Montreal Area (GMA) have been targeted for this study. This area covers approximately 9840 square kilometers and contains a population of roughly 4 million inhabitants (Transport, 2013). Since it is a bilingual region with both French and English speaking populations, the survey was prepared in both languages. In order to decrease the respondent burden, mitigate the implementation cost, and enhance the data quality, a high performance front-end user interface with an elaborated graphic design has been adopted. For more details on the design of the interface, the reader is referred to (Bourbonnais & Morency, 2013).

To minimize the complexity of questions, where respondents had to specify the origin and destination points of their trips and trace the considered routes, geographical map interfaces were adopted. Moreover, an internal validation mechanism was designed to maximize the quality and completeness of the recorded data, and to minimize the data cleaning effort, by reducing participant errors. In this process, several validation criteria were defined for each question, and responses were required to comply with all the criteria in order to be approved and stored in the database. It should be noted that, to advance to a next section, all the responses in that section should be accepted by the internal validation process. In other words, respondents are required to satisfy all the validation criteria of a particular section to be able to advance to the next section. A red exclamation mark appears beside questions that do not meet the required validation criteria, along with a message box explaining the reasons for which the given answers are not acceptable.

Seven types of questions are used in the design of the survey, namely *Dichotomous*, *Text box*, *Select*, *Multi-select*, *Slider*, *Map-point*, and *Map-route* questions. Table 6.2, describes these question types, and provides an illustrative example for each of them.

Table 6.2: Question types used in the survey

Question Type (Data type)	*Description*	*Example*
Dichotomous (Boolean)	Provides two options for a statement to select from. It is also used as a filter or contingency question to determine the next question of the survey.	
Text box (String)	Requires respondents to enter a number, a text, or a combination of both.	
Select (String)	Provides a list of choices, from which respondents can select only one of the several options.	
Multi-select (String)	Allows respondents to choose more than one option from a list.	
Slider (Integer)	Likert scale questions are used to measure attitudes, opinions, perceptions, and levels of agreement with a statement (Likert, 1932). We adopt a *Slider* with a continuous scale to obtain more precise recordings of respondents' views.	
Map-point (Point geometry)	To collect location data (such as origin or destination points), respondents need to pin-point the location on the map.	
Map-route (Route + Point geometry)	Respondents are required to specify their routes by dragging and moving an automatically generated route between the predefined OD pairs.	

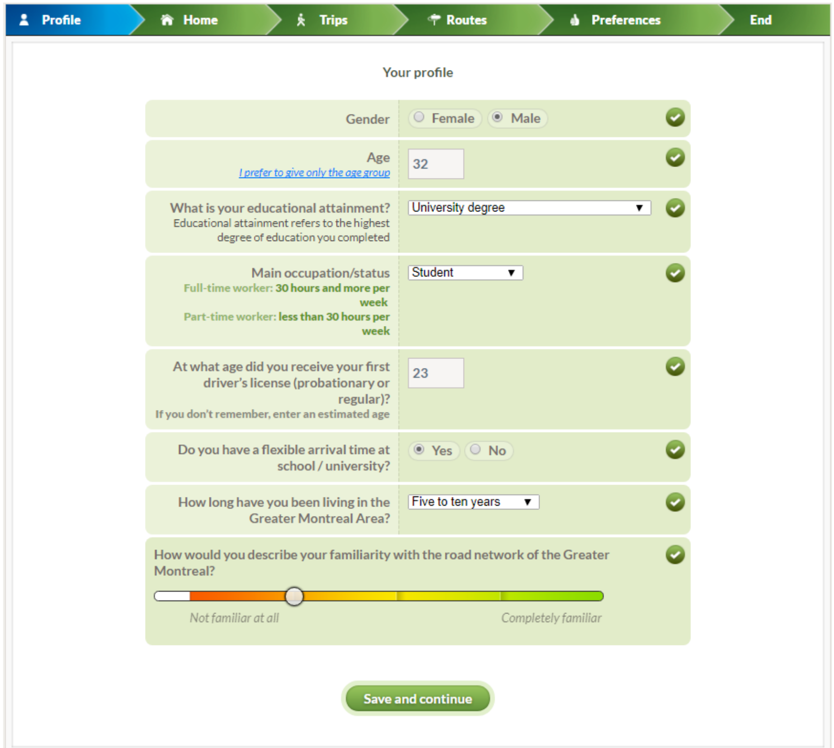
The whole survey is divided into six separate sections, namely *Profile*, *Home*, *Trips*, *Routes*, *Preferences*, and *End*. In the first section, *Profile*, we collect typical sociodemographic data, such as age, gender, educational attainment, type of work, and salary. Collecting these data provides the possibility of comparing the sample population with the reference population (Ory & Mokhtarian, 2005), and to segment the population, based on factors affecting individuals' route choices. Respondents are also asked to specify the duration that they have been living in the GMA and to indicate their general familiarity with its road network.

In the following section, *Home*, participants are required to provide their home address. A geographical map is also provided, which geolocates the specified address. The provided address should be precise enough to be automatically pinpointed on the map. Participants can further adjust the pinpointed location on the map by moving the marker to the exact location. This section also includes questions regarding the household size, the number of cars in the household, the duration of living at the same address, and the familiarity with the road network around the specified address. It is worth mentioning that the exact home address is required to explore factors such as the accessibility to the road network, availability of transit services, and land use specifications.

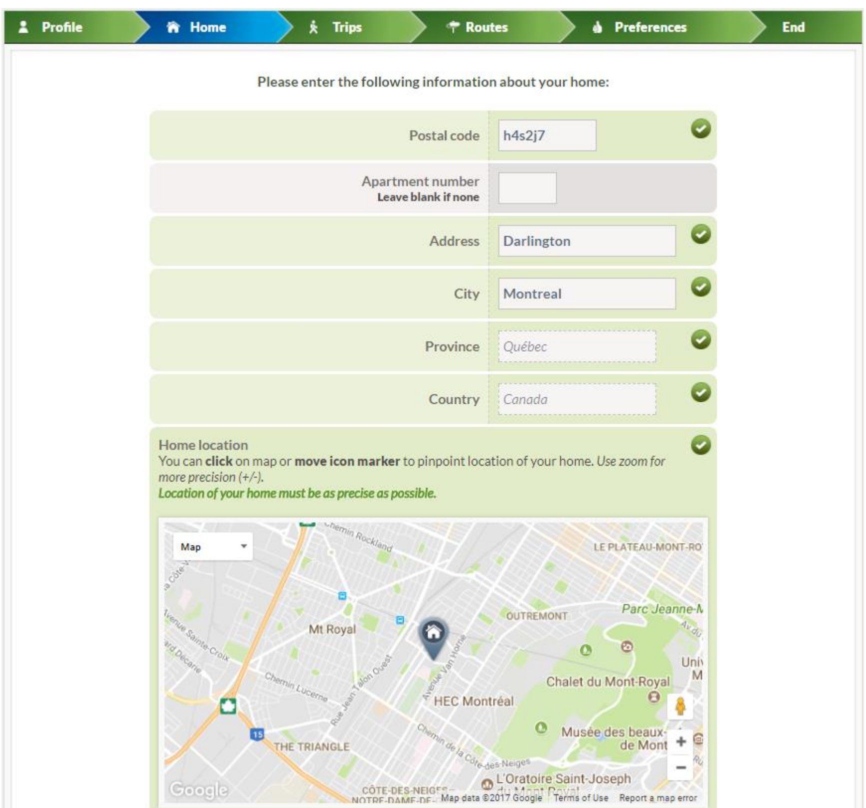
The third section, *Trips*, also consists of a geographical map, on which respondents specify the destination point to which they drive most frequently, such as work places, shopping malls, etc. They are then required to indicate their familiarity with the road network around the specified destination point, and the purpose for making the declared trip. Respondents are also asked to specify why they have used their cars to make the declared trip, as well as to select the five most important factors affecting their route choices for that particular trip. Moreover, they are asked to select from a list, the type of information that they consult prior to their trip and on the way, if any. This section ends by asking respondents to provide the number of route alternatives that they consider for the declared trip. Table 6.3, provides more details regarding questions in the first three sections, and Figure 6.1, demonstrates the web interface for the first two sections.

Table 6.3: List of questions in sections Profile, Home, and Trips

ID	Question Type	Description
First Section: Profile		
101	*Dichotomous*	Gender
102	*Text box / Select*	Age / Age group
103	*Select*	Educational attainment
104	*Select*	Main occupation
105	*Text box*	Age of first driving licence
106	*Dichotomous*	If question ID 104 equals “Worker” ➔ Whether work on the road regularly
107	*Dichotomous*	If question ID 104 equals “Worker” ➔ Whether work at home regularly
108	*Dichotomous*	If question ID 104 equals “Worker” or “Student” ➔ Whether flexible arrival time
109	*Select*	Living time in Montreal
110	Slider	Familiarity with Montreal road network
Second Section: Home		
201	*Text box*	Postal code
202	*Text box*	Apartment number (optional)
203	*Text box*	Street Address
204	*Text box*	City
205	*Map-point*	Home location
206	*Text box*	Household size
207	*Text box*	Household vehicle number
208	*Select*	Living time at the specified address
209	*Slider*	Familiarity with the road network of the neighborhood they live in
Third Section: Trips		
301	*Map-point*	Specify destination point
302	*Slider*	Familiarity with the road network around the destination
303	*Dichotomous*	Is origin home location?
304	*Map-point*	If question ID 303 equals “No” ➔ Specify origin point (if not home location)
305	*Slider*	Familiarity with the road network around the specified origin
306	*Select*	Purpose of the trip
307	*Text box*	Frequency of driving to the specified destination by car (per week)
308	*Multi-select*	Why choose car for this trip?
309	*Multi-select*	Factors affecting route choice to this destination



(a) Section *Profile*



(b) Section *Home*

Figure 6.1: The first two sections of the survey: Profile and Home

Route alternatives are specified in the fourth section, *Routes*. First, an automatically generated route connecting the predefined origin and destination points appears on a geographical map. Then, respondents are required to drag the generated route and adjust it to match their actual considered route. Every time respondents drag the route to a new place on the map, a way point is created on that new location. A minimum of three way points are required for the route to be validated by the internal validation process of the questionnaire. Figure 6.2, demonstrates two alternative routes specified for the same OD pair.

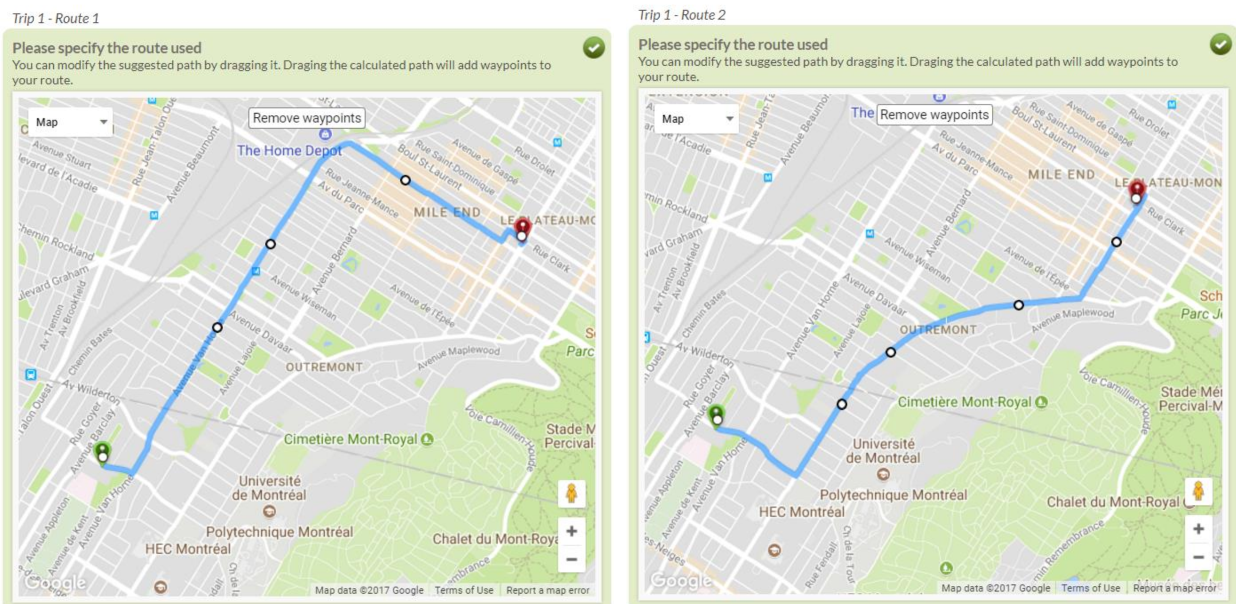


Figure 6.2 : Specifying two route alternatives for the same OD pair in the Routes section

Specified routes are followed by several questions, to gather more details on their main features (see Table 6.4). At first, respondents are required to indicate how frequent they use the declared alternative, on a five level Likert scale ranging from rarely to frequently. Then they provide information regarding the day (i.e. weekdays / week-ends) and the specific time period, during which they start the trip. They also indicate the importance of habit and the effect of weather conditions on their use of the declared route. Moreover, respondents are asked to pinpoint their regular stop (if any) on a map, and specify the amount of toll paid for that particular trip. Finally, drivers' perception regarding the characteristics of the declared alternatives is evaluated based on several factors such as travel time and its reliability, safety, traffic conditions, scenery, pavement quality, and the number of traffic lights.

Table 6.4: List of questions in section Routes

ID	Question Type	Description
401	*Map-route*	Specify considered route
402	*Slider*	Frequency of using the specified route
403	*Select*	Weekdays / Weekend
404	*Select*	Departure time
405	*Text box*	Perceived travel time
406	*Multi-select*	Effect of weather conditions
407	*Dichotomous*	Use tolled route?
408	*Text box*	If question ID 407 equals "Yes" ➔ Amount of toll
409	*Slider*	Perception of safety
410	*Slider*	Perception of scenery
411	*Slider*	Perception of travel time reliability
412	*Slider*	Perception of pavement quality
413	*Slider*	Perception of traffic conditions
414	*Select*	Perception of the number of traffic lights
415	*Slider*	Effect of habit in choosing the specified route
416	*Dichotomous*	Have regular stop?
417	*Map-point*	If question ID 416 equals "Yes" ➔ Specify the location of the regular stop

The fifth section of the survey, entitled *Preferences*, focuses on behavioural and attitudinal variables affecting drivers' route choice behaviours. A list of different statements is provided to respondents, who were asked to specify their level of agreement with each statement on a five-point Likert scale ranging from total agreement to total disagreement. These statements are based on psychometric indicators and on some behavioural assumptions on drivers' attitudes, to reveal the most important latent variables affecting drivers' route choice behaviours (Atasoy, Glerum, & Bierlaire, 2013; Ory & Mokhtarian, 2005; Vredin Johansson, Heldt, & Johansson, 2006). Table 6.5, presents the statements included in this survey, and Figure 6.3, illustrates the first few statements.

The survey ends with few optional questions in the final section, *End* (see Table 6.6). First, respondents are asked to provide their household's gross income level. Then, they are asked to provide their e-mail address if they desire to participate in other transportation surveys. Finally, respondents can provide their comments and opinions regarding the survey in a blank box.

To evaluate the simplicity and clarity of the questions, to assess the accuracy of the provided directions on how to complete the survey, and to detect the weaknesses of the designed interface, graduate students of the Transportation Research Group of Polytechnique Montreal took part in a pilot test in February 2017. The revised version of the survey was launched in March 2017, and data was collected over a period of three months.

Table 6.5 : List of statements in the fifth section of the survey, Preferences

ID	Statement
501	Driving to my destination, I prefer to take freeways, whenever I have access to them.
502	Driving to my destination, I prefer to take local routes, even when freeways are available.
503	The pavement quality is an important factor in my route choice.
504	I always look for shortcuts to minimize the travelled distance.
505	I do my best to avoid traffic lights.
506	Minimizing the travel time is my principal goal while choosing my route.
507	I prefer taking a longer route with a fluid traffic flow rather than being stuck in traffic in a shorter one.
508	I have the tendency to follow the same route over and over.
509	I have the tendency to try new routes.
510	I tend to avoid routes with narrow lanes.
511	I prefer to take routes with higher speed limits.
512	I am not comfortable driving next to trucks and I try to avoid them.
513	I prefer to choose a more beautiful and scenic route, even if it takes longer to get to work.
514	I prefer to take tolled routes because they are less congested and much faster.
515	I inform myself about road construction sites to avoid them.
516	I have the tendency to avoid turns and take the most direct route to get to work.
517	I have a good sense of direction and I can easily find my way in a road network.
518	When I'm informed by radio or variable message signs, of an accident causing traffic jam on my route, I change my itinerary and choose an alternative route to avoid the congestion.
519	I can easily remember a route which I took once.
520	I use landmarks to remember a route that I took once.
521	I prefer to choose a route which has a more reliable travel time even if it takes me more time.
522	I take the route suggested by Google Maps (or other route planners).

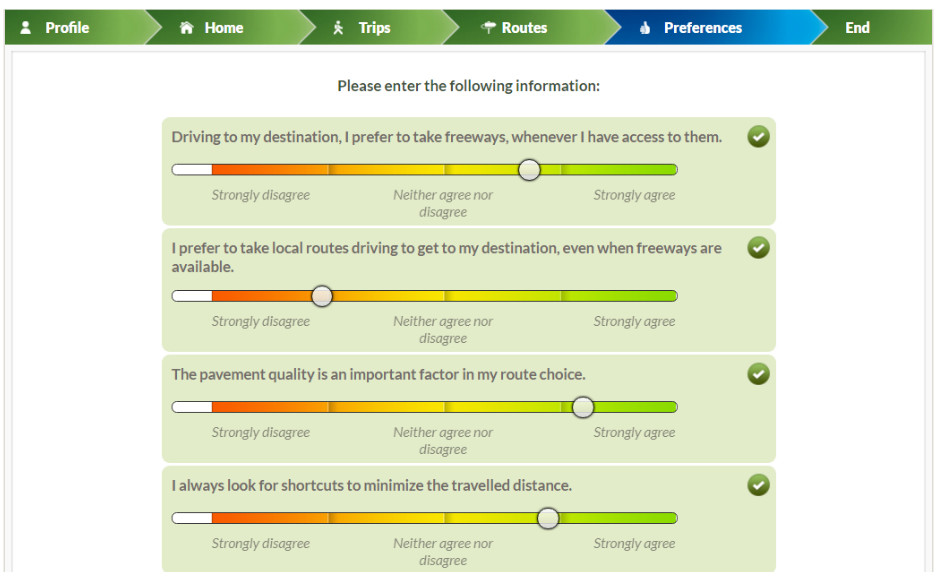


Figure 6.3: Fifth section of the survey, Preferences

Table 6.6: List of questions in section End

ID	Question Type	Description
601	*Select*	Household gross income level
602	*Dichotomous*	Would like to participate in other mobility surveys
603	*Text box*	If question ID 602 equals "Yes" ➔ Put e-mail address
604	*Text box*	General comments on the interview

6.4 Survey Response Behaviour

By the end of the three-month data collection period, 843 individuals started the survey from which 539 (64 %) completed it, while the remaining 304 (36 %) dropped out at various points of the survey. In this section, we present the survey recruitment methods and the obtained response rates. We also investigate participants' characteristics and their response behaviours. Finally, we explore participants who dropped out of the survey before finishing it.

6.4.1 Recruitment and response rates

To be eligible to take part in the survey, participants were required to reside in the GMA and drive to at least one specific destination in this area. These criteria were clearly outlined in several occasions, including the invitation letter, the starting page of the survey, and the informed consent form, which was mandatory to read and accept before starting the survey. To disseminate the survey, three target groups were identified: i) graduate students, postdocs, faculty members, and staff of Polytechnique Montreal, ii) users of social media, such as Facebook, LinkedIn, etc., and iii) volunteer participants who previously agreed to participate in surveys conducted by the Mobility Chair of Polytechnique Montreal and provided their e-mail addresses. Figure 6.4, illustrates the number of completed surveys in each day for the total period of data collection, as well as the recruitment methods that have been adopted.

To reduce the number of simultaneous respondents, and hence, the load on the server hosting the web survey, volunteers' e-mail addresses have been divided into five separate lists with smaller numbers of e-mail addresses. A total of 4000 volunteers were contacted on different occasions, and a recall e-mail has been sent to those who have received the first invitation letter, few days later. The number of e-mails sent, delivered, opened by the receiver, as well as the number of recipients who clicked on the survey link (enclosed in the invitation letter) are reported in Table 6.7. Out of the 95 % of recipients who received the first invitation letter, 45.1 % opened the e-mail and 12.7 % clicked on the survey link. However, for the recall e-mail, these statistics were 98.6 %, 45.3 %, and 10.1 %, respectively.

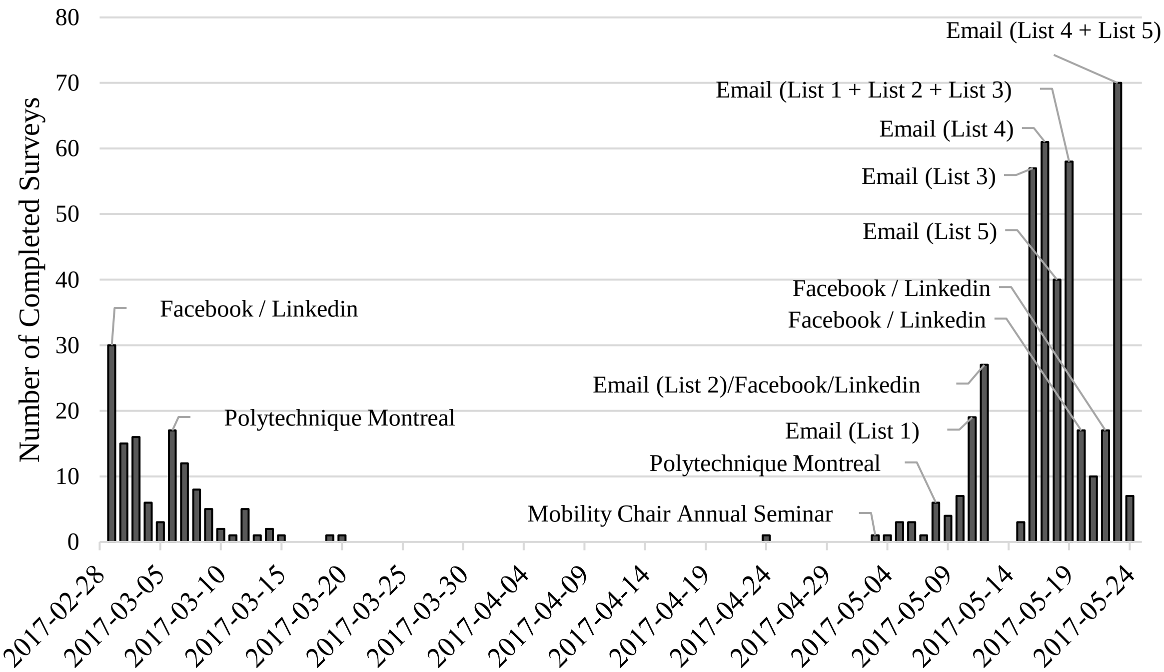


Figure 6.4: Recruitment method and response rates during the data collection period

Table 6.7: Invitation letters sent to volunteers' lists

Date	Time of Day	List Name	Sent	Delivered	Opened	Clicked
<u>*Invitation E-mail*</u>						
11-May	9:00 AM	List 1	500	433 (86.6 %)	170 (34.0 %)	35 (7.0 %)
12-May	9:00 AM	List 2	500	478 (95.6 %)	260 (52.0 %)	63 (12.6 %)
16-May	9:00 AM	List 3	1000	959 (95.9 %)	528 (52.8 %)	140 (14.0 %)
17-May	10:00 AM	List 4	1000	952 (95.2 %)	424 (42.4 %)	136 (13.6 %)
18-May	10:00 AM	List 5	1000	955 (95.5 %)	422 (42.2 %)	132 (13.2 %)
<u>*Recall E-mail*</u>						
19-May	11:00 AM	List 1+ List 2	911	903 (99.2 %)	457 (50.2 %)	100 (11.0 %)
19-May	11:00 AM	List 3	959	952 (99.3 %)	518 (54.1 %)	102 (10.7 %)
23-May	9:00 AM	List 4	952	942 (99.0 %)	369 (38.8 %)	105 (11.1 %)
23-May	9:00 AM	List 5	955	929 (97.3 %)	366 (38.4 %)	98 (10.3 %)

Figure 6.5, illustrates the percentage of completed surveys per hour of the day. Given that invitation e-mails were sent between 9:00 AM and 11:00 AM (see Table 6.7), higher response rates between these hours were expected. The completion rate decreases throughout the afternoon, increases slightly around 9:00 PM to 10:00 PM, and reaches its minimum overnight.

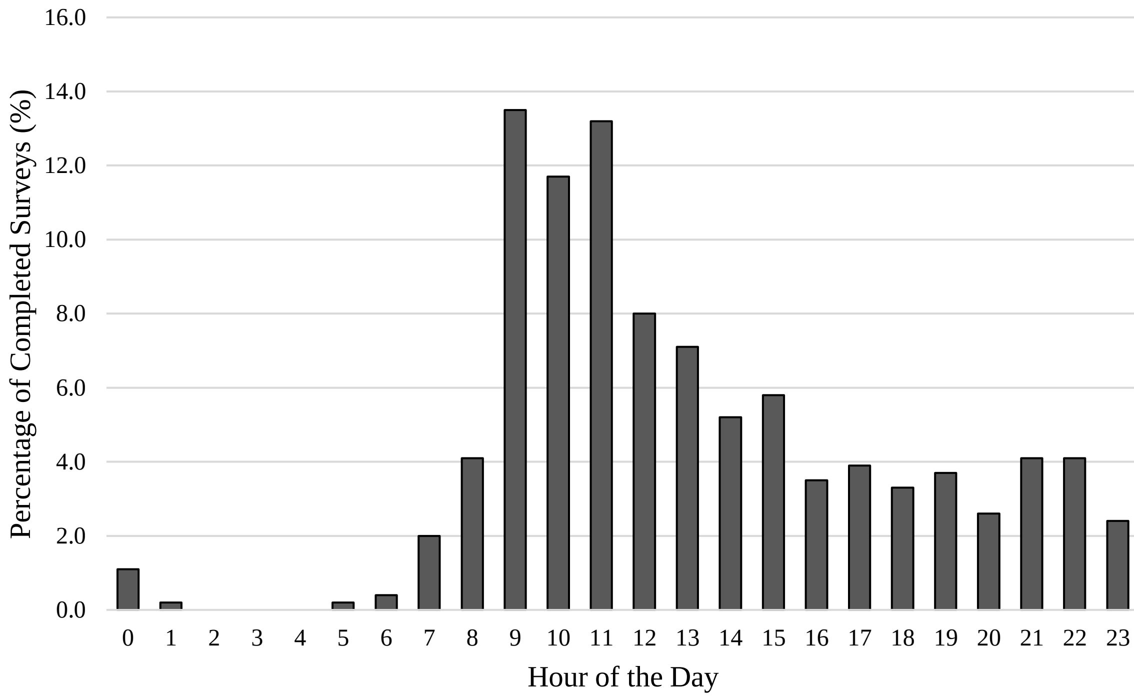


Figure 6.5: Percentage of completed surveys in each hour of the day for the whole data collection period

6.4.2 Participants Characteristics

Table 6.8, illustrates sociodemographic and socioeconomic characteristics of the 539 respondents who have completed the survey. It should be noted that the sample includes mainly young and middle aged full time workers with a university level of education. This may partly be because the survey was also disseminated among scholars, faculty members, and staff of Polytechnique Montreal. Moreover, the prevalence of young participants explains to some degree the higher frequency of smaller households.

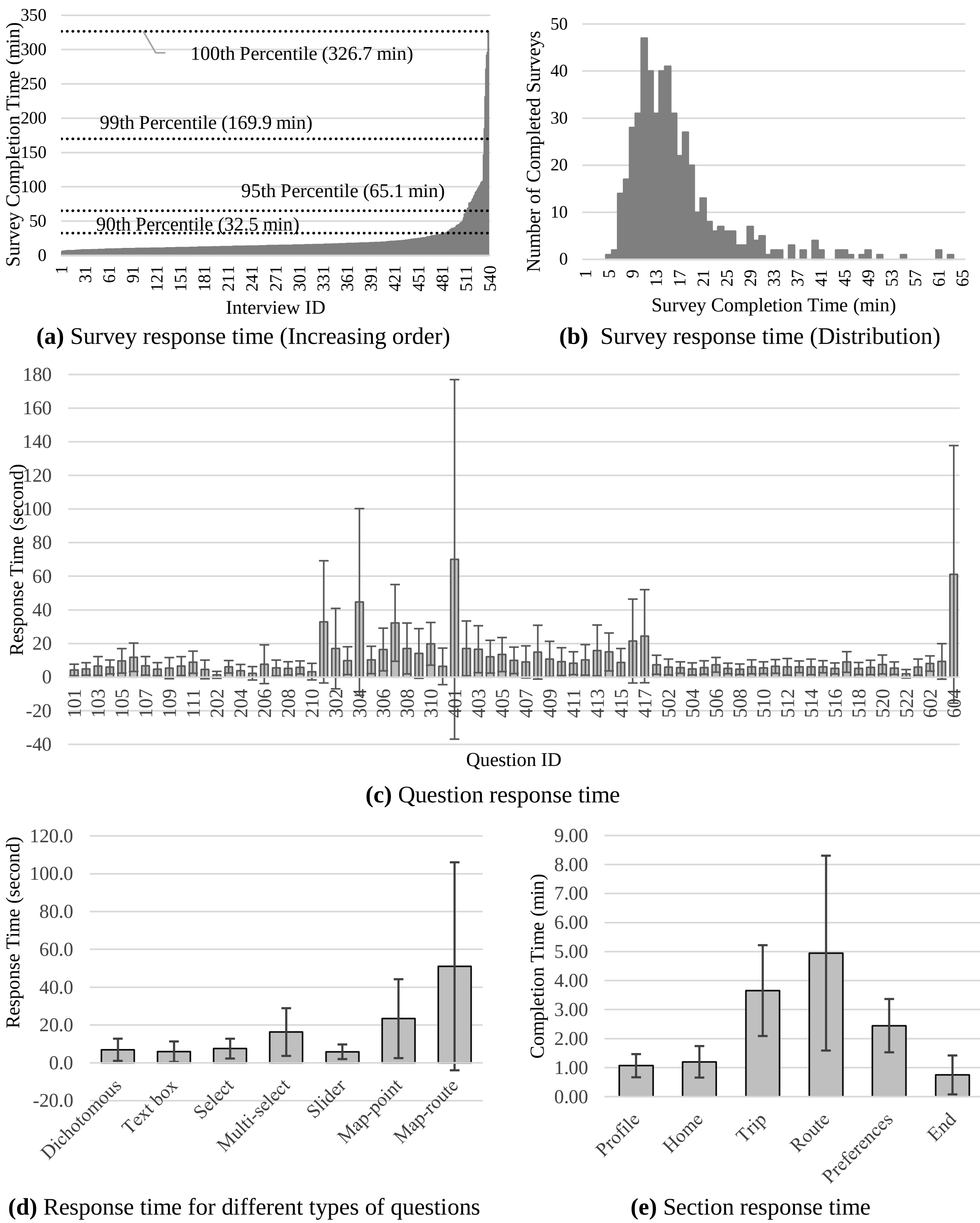
Table 6.8: Characteristics of the Survey Participants

Variable Categories	N	%	Variable Categories	N	%
Gender			*Household size*		
Male	306	56.7	1	100	18.5
Female	234	43.3	2	218	40.4
Age (years old)			3	101	18.7
Young (15 to 39)	306	56.7	4	88	16.3
Middle age (40 to 59)	195	36.1	+5	33	6.1
Old (more than 60)	39	7.2			
Occupation			*Income (Thousand CAD per capita)*		
Full time worker	393	72.8	< 30	152	28.2
Partial time worker	38	7.0	> 30 and < 60	202	37.4
Student	69	12.8	> 60 and < 90	75	13.9
Retired	25	4.6	> 90 and < 120	12	2.2
House-wife/husband	6	1.1	Not declared	99	18.3
Other	9	1.7			
Education			*Household car number*		
None	0	0.0	0	122	22.6
Less than university	62	11.5	1	290	53.7
University	472	87.4	2	106	19.6
Other	6	1.1	+3	22	4.1

6.4.3 Response Behaviour

Usage information showed that 61% of participants used Windows devices to complete the survey, while Macs (20 %), IOS (12%), Android (5%), Linux (1%), and Chrome OS (1%) accounted for the remaining 39%. Moreover, information on the variety of web browser illustrates that Chrome (59%), Firefox (18%), and Safari (15%) account for around 92% of the completed surveys, while the remaining 8% have been completed on Internet Explorer, Microsoft Edge, and Opera. These statistics emphasize the importance of making the interface friendly and easy to use for a wide range of devices and browsers to increase the response rate of a survey.

Figure 6.6(a), illustrates survey completion times (in minutes) in an increasing order. The completion time of the survey is expressed as the summation of the completion time of all the sections for each interview. It can also be thought of as the difference between the starting time and ending time of the survey excluding the time that respondents had left the survey platform. Considering a 95th percentile threshold, the average and maximum completion time of the survey are found to be 16.1 and 65.1 minutes, respectively. The distribution of the survey completion time is illustrated in Figure 6.6(b).



(a) Survey response time (Increasing order) **(b)** Survey response time (Distribution)

(c) Question response time

(d) Response time for different types of questions **(e)** Section response time

Figure 6.6: Response times (Vertical bars represent the standard deviation)

Considering a 95th percentile threshold, the average response time for different questions, question types, and sections are illustrated in Figure 6.6(c), Figure 6.6(d), and Figure 6.6(e), respectively. It can be noted that question ID 401 has the highest response time and variation. Given the complexity of the question, which involves a geographical map and requires respondents to drag and adjust a suggested route, the high response time of this question is not surprising. It is noticed that questions involving a geographical map (i.e. *Map-route* and *Map-point* types of questions) have longer response times, while *Dichotomous*, *Text-box*, *Select*, and *Slider* questions require shorter response times (Figure 6.6(d)). Consequently, sections including more geographical maps (*Trips*, and *Routes*) have longer response times (Figure 6.6(e)).

6.4.4 Dropouts

The number of dropouts per section and question is illustrated in Figure 6.7. Most of the dropouts occur in the *Trips* section (46.1 %). This may mostly be because some respondents started the questionnaire without satisfying the required participation criteria, i.e. residing and driving in the GMA. In this section, respondents are asked specific questions regarding a destination to which they drive frequently (such as questions 301, 302, and 306). It may be at this point of the survey that they realize that they are not fit to continue the survey. We received several e-mails, Facebook messages, and survey comments supporting the claim that some respondents failed to pay sufficient attention to the participation criteria. As mentioned before, these criteria were repeatedly mentioned in the invitation letter, survey starting page, as well as the informed consent form. The same argument stands for the higher rate of dropouts in question 106 (in section one), in which respondents are required to declare the age at which they got their driving licence.

The second section with the highest dropout rate is the *Routes* section, in which the first question (i.e. specifying the considered route) has the highest dropout rate of the section. Considering that detailed instructions were provided on how to specify routes on a geographical map, both in the introductory page of the *Routes* section as well as on top of question 401, the high rate of dropouts may be related to the innate complexity of working with geographical maps and the longer response time required for this question.

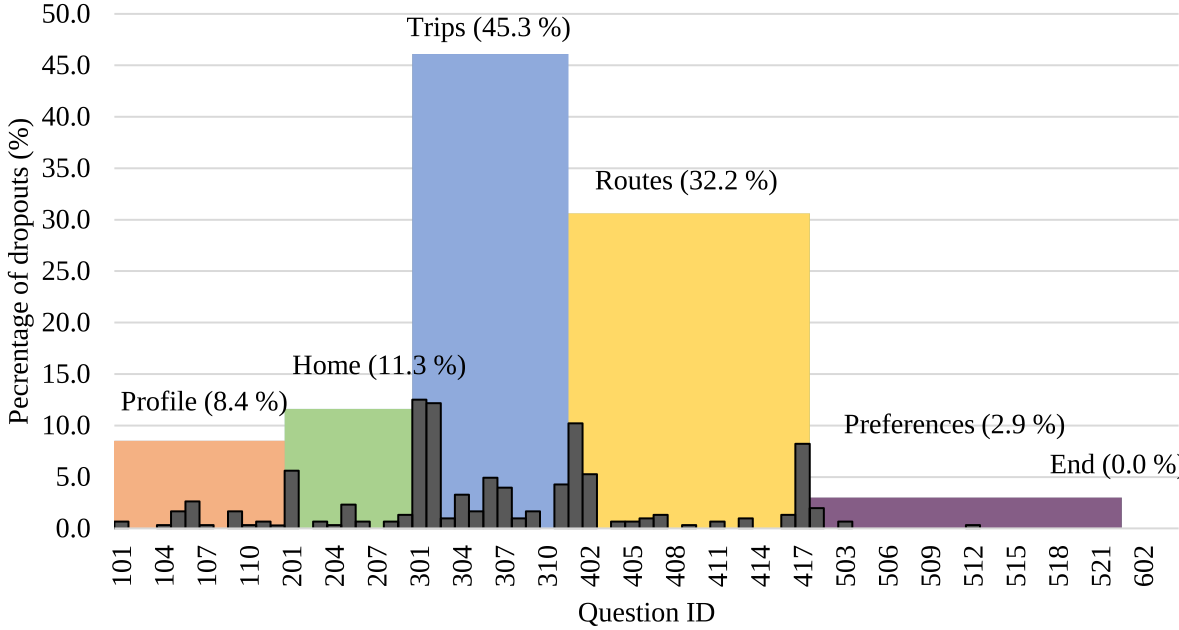


Figure 6.7: Percentage of dropouts from the survey per section and question

Interestingly, we notice that the first question of every section (except for the first and last sections) has the highest rate of dropouts in that respective section. This indicates that reducing the number of new sections may decrease the total number of dropouts in the survey.

To compare the dropout rates of different question types, the total number of dropouts for each question type is divided by the number of recurrence of that particular question type in the whole survey (see Table 6.9). For instance, 20 dropouts occurred within the 8 recurrences of Dichotomous questions, resulting a dropout rate of 2.5 (i.e. 20/8). Results demonstrate that *Map-route* and *Map-point* question types induce higher dropouts compared to other types of questions. Moreover, the effect of *Multi-select* questions is found to be more pronounced than *Select* questions. It can also be noted that *Slider* and *Dichotomous* questions have the least effects on the number of dropouts.

Table 6.9: Dropout rates for different types of questions

Question Type	Dropouts	Recurrence	Dropout rate
Dichotomous	20	8	2.5
Text box	47	13	3.6
Select	30	10	3.0
Multi-select	23	5	4.6
Slider	68	33	2.1
Map-point	79	4	19.8
Map-route	31	1	31.0

6.4.5 Survey Comments

Respondents were asked to provide their general comments, in the last question of the survey. It should be noted that only respondents who advanced to the sixth section of the survey could leave a comment and those who abandoned the survey before that section could not access the comment box to leave their comments. A total number of 149 respondents left comments, most of which were generally positive and encouraging, commending the objectives, the question design and the appearance of the survey. Few respondents, however, criticized the length and complexity of the questionnaire and the hardships of specifying a route trajectory on geographical map. They also reported some technical issues regarding some browsers and occasional difficulties with phone interfaces. We have also received few messages and e-mails from respondents concerned about privacy aspects, most of whom left the survey in the second section. Interestingly, a significant number of respondents who left comments were not happy about the exclusive focus of the survey on drivers' route choices, and asked for a more comprehensive travel survey, considering other modes of transportation such as public transit, walk, and bike.

6.5 Conclusion

The increasing application of advanced choice models, reflecting the stochasticity of individuals' preferences and the complex nature of human decision-making behaviour, requires enhanced data collection methods collecting detailed data without significantly increasing respondent burden. This paper details the development and deployment of a general survey framework for route choice studies with three main objectives: i) to observe drivers revealed route choices, ii) to identify important factors including behavioural and attitudinal factors affecting them, and iii) to observed and characterize drivers' consideration sets of route alternatives.

A web-based survey has been designed to provide researchers with a rich dataset, based on which they can produce reliable behavioural models. The adopted graphical interface is expected to augment response precision and to reduce the burden of declaring all the considered alternatives. Moreover, the analyst obtains the exact trajectories considered for each trip and will not face the challenges and uncertainties associated with GPS datasets such as trip extraction, map-matching, and path inference. In short, the analyst will be able to investigate more closely some major challenges facing route choice modelling, such as the definition of an alternative route and how it is perceived by drivers, the characteristics of a considered set of route alternatives, and the role of different attributes (observable and latent) in route choice decisions.

Considering the high number of questions included in the survey (74 questions), the variety of question types (7 types), and the spectrum of collected data (i.e. sociodemographic, socioeconomic, revealed preference route choices, considered sets of route alternatives, decision maker's perceptions, and behavioural traits), the overall survey completion percentage of 64 % suggest a successful implementation of the survey framework. An internal validation system has been applied to minimize participants' errors and maximize the completeness of survey responses. As a result, a small number of interviews were discarded (26 out of 539, 4.8 %), which indicate the high quality of the collected data. The principal reasons for exclusion were twofold: first, living or driving outside the study area, and second, failing to specify a logically sound route between the predefined OD pairs. To increase the quality of the final dataset, unusually short or long response times can be used as proxy indicators to identify measurement errors (Couper & Kreuter, 2013).

Despite the successful application of the proposed survey framework, the authors acknowledge its limitations and the uncertainties associated with certain responses. For instance, like HTS, in which trip under-reporting is an undeniable issue, the possibility of under-reporting the number of considered route alternatives for the declared trip is recognized. Moreover, it is not straightforward to ensure a representative sample of the population using web-based surveys. For instance, older individuals may have limited access to the internet, or may lack the technical knowledge to answer to certain types of questions (Bourbonnais & Morency, 2013).Further research could include a comparison of the proposed framework with other route choice data collection frameworks, with different types of questions and various lengths to better evaluation of the performance and data quality of different frameworks. Also, the completion time can be used as an indicator of respondent burden (Greaves et al., 2015), and the effect of different completion times can be studied on dropout rates. Another possible extension can be the integration of the proposed survey framework with smartphones and GPS devices to compare declared with actual route choices.

Acknowledgement

The authors would like to express their gratitude to the partners of the Mobilité Research Chair of Polytechnique Montréal for their financial support and their help in the dissemination of the survey.

References

Abdel-Aty, M., Jovanis, P., 1997. Using GIS Capabilities to Improve The Understanding Of Route Choice Behavior, Transportation Research Board, Washington D.C.

Alizadeh, H., Farooq, B., Morency, C., Saunier, N., 2017. On the role of bridges as anchor points in route choice modelling. Transportation.

Atasoy, B., Glerum, A., Bierlaire, M., 2013. Attitudes towards mode choice in Switzerland. disP - The Planning Review 49, 101-117.

Ben-Akiva, M., Walker, J., Bernardino, A.T., Gopinath, D.A., Morikawa, T., Polydoropoulou, A., 2002. Integration of choice and latent variable models. Perpetual motion: Travel behaviour research opportunities and application challenges, 431-470.

Ben-Elia, E., Shiftan, Y., 2010. Which road do I take? A learning-based model of route-choice behavior with real-time information. Transportation Research Part A: Policy and Practice 44, 249-264.

Bliemer, M., Bovy, P., 2008. Impact of route choice set on route choice probabilities. Transportation Research Record: Journal of the Transportation Research Board, 10-19.

Bourbonnais, P.-L., Morency, C., 2013. Web-Based Travel Survey: A Demo, in: Johanna Zmud, M.L.-G., Marcela Munizaga, Juan Antonio Carrasco (Ed.), Transport Survey Methods: Best Practice for Decision Making, pp. 207-224.

Bradley, M., 2013. Workshop Synthesis: Measuring the Influence of Attitudes and Perceptions, Transport Survey Methods: Best Practice for Decision Making. Emerald Group Publishing Limited, pp. 643-646.

Cascetta, E., Russo, F., Viola, F.A., Vitetta, A., 2002. A model of route perception in urban road networks. Transportation Research Part B: Methodological 36, 577-592.

Chen, J., 2013. Modeling route choice behavior using smartphone data, FACULTÉ DE L'ENVIRONNEMENT NATUREL, ARCHITECTURAL ET CONSTRUIT. ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE, Suisse.

Cools, M., Moons, E., Wets, G., 2009. Assessing the Impact of Public Holidays on Travel Time Expenditure Differentiation by Trip Motive. NATL ACAD SCIENCES.

Couper, M.P., Kreuter, F., 2013. Using paradata to explore item level response times in surveys. Journal of the Royal Statistical Society: Series A (Statistics in Society) 176, 271-286.

Dalumpines, R., Scott, D.M., 2017. Determinants of route choice behavior: A comparison of shop versus work trips using the Potential Path Area - Gateway (PPAG) algorithm and Path-Size Logit. Journal of Transport Geography 59, 59-68.

Dhakar, N.S., 2012. Route Choice Modeling Using GPS Data. University of Florida.

Gan, H., Chen, S., 2013. Why Do Drivers Change Routes? Impact of Graphical Route Information Panels. ITE Journal.

Gärling, T., Gillholm, R., Gärling, A., 1998. Reintroducing attitude theory in travel behavior research: The validity of an interactive interview procedure to predict car use. Transportation 25, 129-146.

Geda, A., 2014. Effect of Choice Set Composition on Route Choice Models. University of Florida.

Greaves, S., Ellison, A., Ellison, R., Rance, D., Standen, C., Rissel, C., Crane, M., 2015. A web-based diary and companion smartphone app for travel/activity surveys. Transportation Research Procedia 11, 297-310.

Habib, K.N., Morency, C., Trépanier, M., Salem, S., 2013. Application of an independent availability logit model (IAL) for route choice modelling: Considering bridge choice as a key determinant of selected routes for commuting in Montreal. Journal of Choice Modelling 9, 14-26.

Hess, S., Daly, A.J., 2010. Choice Modelling: The State-of-the-Art and the State-of-Practice: Proceedings from the Inaugural International Choice Modelling Conference: The State-of-the-art and the State-of-practice: Proceedings from the Inaugural International Choice Modelling Conference. Emerald Group Publishing.

Hess, S., Quddus, M., Rieser-Schüssler, N., Daly, A., 2015. Developing advanced route choice models for heavy goods vehicles using GPS data. Transportation Research Part E: Logistics and Transportation Review 77, 29-44.

Hoogendoorn-Lanser, S., van Ness, R., Bovy, P.H., 2005. Path Size and Overlap in Multi-modal Transport Networks, Transportation and Traffic Theory. Flow, Dynamics and Human Interaction. 16th International Symposium on Transportation and Traffic Theory.

Iida, Y., Uno, N., Yamada, T., 1994. Experimental analysis approach to analyze dynamic route choice behavior of driver with travel time information, Vehicle Navigation and Information Systems Conference, 1994. Proceedings., 1994, pp. 377-382.

Jan, O., Horowitz, A., Peng, Z.-R., 2000. Using global positioning system data to understand variations in path choice. Transportation Research Record: Journal of the Transportation Research Board, 37-44.

Jou, R.-C., Yeh, Y.-C., 2013. Freeway passenger car drivers' travel choice behaviour in a distance-based toll system. Transport Policy 27, 11-19.

Kamargianni, M., Dubey, S., Polydoropoulou, A., Bhat, C., 2015. Investigating the subjective and objective factors influencing teenagers’ school travel mode choice – An integrated choice and latent variable model. Transportation Research Part A: Policy and Practice 78, 473-488.

Kaplan, S., Prato, C.G., 2012. Closing the gap between behavior and models in route choice: The role of spatiotemporal constraints and latent traits in choice set formation. Transportation Research Part F: traffic psychology and behaviour 15, 9-24.

Koller-Matschke, I., Belzner, H., Glas, R., 2013. Traffic Information and Individual Driver Behaviour, Proceedings of the FISITA 2012 World Automotive Congress. Springer Berlin Heidelberg, pp. 179-185.

Lai, X., Bierlaire, M., 2015. Specification of the cross-nested logit model with sampling of alternatives for route choice models. Transportation Research Part B: Methodological 80, 220-234.

Likert, R., 1932. A technique for the measurement of attitudes. Archives of psychology.

Manley, E.J., Addison, J.D., Cheng, T., 2015. Shortest path or anchor-based route choice: a large-scale empirical analysis of minicab routing in London. Journal of Transport Geography 43, 123-139.

Manski, C.F., 1977. The structure of random utility models. Theory and decision 8, 229-254.

McFadden, D., 1999. Rationality for Economists? Journal of Risk and Uncertainty 19, 73-105.

Muñoz, B., Monzon, A., López, E., 2016. Transition to a cyclable city: Latent variables affecting bicycle commuting. Transportation Research Part A: Policy and Practice 84, 4-17.

Nadine Schuessler, I., Axhausen, K.W., 2009. Accounting for Route Overlap in Urban and Sub-urban Route Choice Decisions Derived from GPS Observations, 12th International conference on travel behavior research, Jaipur, India.

Ory, D.T., Mokhtarian, P.L., 2005. Don’t work, work at home, or commute? Discrete choice models of the decision for San Francisco Bay Area residents.

Papinski, D., 2011. Investigating Route Choice Decisions Using GPS and Prompted-recall Diary Data, School of Graduate Studies. McMaster University.
Papinski, D., Scott, D.M., Doherty, S.T., 2009. Exploring the route choice decision-making process: A comparison of planned and observed routes obtained using person-based GPS. Transportation Research Part F: Traffic Psychology and Behaviour 12, 347-358.
Parkany, E., Du, J., Aultman-Hall, L., Gallagher, R., 2006. Modeling Stated and Revealed Route Choice: Consideration of Consistency, Diversion, and Attitudinal Variables. Transportation Research Record: Journal of the Transportation Research Board 1985, 29-39.
Peeta, S., Ramos, J., Pasupathy, R., 2000. Content of Variable Message Signs and On-Line Driver Behavior. Transportation Research Record: Journal of the Transportation Research Board 1725, 102-108.
Peters, T., Adamowicz, W.L., Boxall, P.C., 1995. Influence of choice set considerations in modeling the benefits from improved water quality. Water Resources Research 31, 1781-1787.
Prato, C., Bekhor, S., 2006. Applying branch-and-bound technique to route choice set generation. Transportation Research Record: Journal of the Transportation Research Board, 19-28.
Prato, C., Bekhor, S., 2007. Modeling route choice behavior: how relevant is the composition of choice set? Transportation Research Record: Journal of the Transportation Research Board, 64-73.
Prato, C.G., Bekhor, S., Pronello, C., 2012. Latent variables and route choice behavior. Transportation 39, 299-319.
Ramaekers, K., Reumers, S., Wets, G., Cools, M., 2013. Modelling Route Choice Decisions of Car Travellers Using Combined GPS and Diary Data. Netw Spat Econ 13, 351-372.
Sarkar, P.P., Mallikarjuna, C., 2017. Effect of perception and attitudinal variables on mode choice behavior: A case study of Indian city, Agartala. Travel Behaviour and Society.
Schlaich, J., 2010. Analyzing route choice behavior with mobile phone trajectories. Transportation Research Record: Journal of the Transportation Research Board, 78-85.
Srinivasan, S., Dhakar, N.S., 2013. Route-choice Modeling using GPS-based Travel Surveys.
Swait, J., Ben-Akiva, M., 1987. Empirical test of a constrained choice discrete model: mode choice in Sao Paulo, Brazil. Transportation Research Part B: Methodological 21, 103-115.
Tawfik, A.M., Rakha, H., 2013. A Latent Class Choice Model of Heterogeneous Drivers Route Choice Behavior Based on a Real-World Experiment, 20th International Symposium on Transportation and Traffic Theory, Noordwijk, Netherlands.
Tawfik, A.M., Rakha, H.A., Miller, S.D., 2010. An experimental exploration of route choice: Identifying drivers choices and choice patterns, and capturing network evolution, Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference, pp. 1005-1012.
Transport, A.M.d., 2013. Enquête Origine-Destination 2013: La mobilité des personnes dans la région de Montréal-FAITS SAILLANTS, in: Montreal, A.M.d.T. (Ed.), Montreal.
Vacca, A., Meloni, I., 2014. Understanding route choice behavior: an analysis using GPS based data, Transport Research Arena (TRA) 5th Conference: Transport Solutions from Research to Deployment.
Vredin Johansson, M., Heldt, T., Johansson, P., 2006. The effects of attitudes and personality traits on mode choice. Transportation Research Part A: Policy and Practice 40, 507-525.
Walker, J.L., 2001. Extended discrete choice models: integrated framework, flexible error structures, and latent variables. Massachusetts Institute of Technology.

CHAPTER 7 ARTICLE 4: FREQUENT VERSUS OCCASIONAL DRIVERS: A HYBRID ROUTE CHOICE MODEL

Hamzeh Alizadeh[1*], Bilal Farooq[2], Catherine Morency[1], Nicolas Saunier[1]

Under review (revision requested) in *Transportation Research Part F: Traffic Psychology and Behaviour*

Submission Date: 5 / February / 2018

[1] Department of Civil, Geological and Mining Engineering, Polytechnique Montréal, C.P. 6079, succ. Centre-Ville Montréal, Canada

[2] Department of Civil Engineering, Ryerson University, 350 Victoria Street, Toronto, Canada

[*] Corresponding author

Abstract

Previous route choice studies have mostly focused on the effect of observable factors, such as route attributes and socio-economic characteristics, on individuals' decisions. However, route choice decisions might not be exclusively dependent on these observable variables, but also on latent variables, which cannot be directly observed and measured. Also, the latent behavioural heterogeneity among the population has mostly been ignored by assuming that all the individuals in the sample population have similar attitudes, perceptions, and lifestyles. In this paper, we present a comprehensive framework to explicitly incorporate latent behavioural constructs as well as segment heterogeneity based on a probabilistic segmentation of the population. We apply the proposed framework to compare the route choice behaviour of frequent versus occasional drivers. An Integrated Choice and Latent Variable (ICLV) model is used to bring in the role of the underlying behavioural constructs, while a Latent Class (LC) model accounts for taste heterogeneity across the two segments of our sample population. An Extended Path-Size Logit (EPSL) model is adopted as the choice model component and a Metropolis-Hastings (MH) based algorithm is used to generate route alternatives. Data is collected through a web-based survey designed to collect behavioural data on drivers' route choices, using psychometric indicators and behavioural questions on respondents' perceptions and attitudes. Results confirmed that the inclusion of latent variables and latent heterogeneity across population segments significantly improve the explanatory power of the choice model.

Keywords: Route Choice Modelling, Behavioural Heterogeneity, Latent Class Models, Integrated Choice and Latent Variable Models, Revealed Preference Data, Web-Based Survey

7.1 Introduction

How drivers decide upon which route to take? What are the most important factors affecting their choices? How would they react if changes were made to the road network? These are among the questions that route choice models aim to shed some light on. Such models investigate the process of route selection by individuals making a trip between predefined origin and destination (OD) pairs, and assign a selection probability to paths connecting them (Bierlaire & Frejinger, 2008; Prato, 2009b). They are at the core of traffic assignment models and play a crucial role in transportation planning, to forecast the traffic flow, design new transportation infrastructures, and investigate new policies. Therefore, it is of utmost importance that we understand drivers' route choice behaviours and factors affecting them (Li, Miwa, Morikawa, & Liu, 2016; Prato, 2009b).

The complex behaviour of route choice has been studied for decades and different forms of models have been developed to properly represent it (e.g., prospect theory (Gao et al., 2010), cumulative prospect theory (Xu et al., 2011), neural network (Dougherty, 1995), fuzzy preference (Ridwan, 2004), etc.). Random utility discrete choice models are among the most frequently used approaches to model, analyze and understand these behaviours (Prato, 2009b; Walker, 2001). Following the two-stage choice process proposed by Manski (1977), decision makers reduce the universal set of possible alternatives to a considered set of route alternatives in the first stage, from which they make their final choice in the second stage. These models have roots in the consumer theory of microeconomics and assume that individuals, as rational decision makers, choose the most preferred alternative amongst a finite set of alternatives. This preference is expressed through a specific continuous function, called the utility function, which consists of a deterministic component accounting for observable factors (such as route length, travel time, number of turns, etc.), and a stochastic component, i.e. random term, accounting for unobservable attributes (such as attitudes, preferences, etc.). In other words, observed choices are manifestations of decision makers' preferences, expressed by alternative specific utility functions. Estimation of these models results in selection probabilities for each alternative, which can then be used to predict the behaviours of decision makers (Frejinger et al., 2009; Manski, 1977; Walker, 2001).

Previous route choice studies have mostly focused on the effects of observable factors, such as route attributes and socio-economic characteristics (Dalumpines & Scott, 2017a; Jan et al., 2000). According to (Jan et al., 2000), factors affecting route choice decisions can be classified into four categories including *travellers' attributes* (such as age, gender, education, income, etc.), *route attributes* (such as traffic conditions, speed limits, number of turns, pavement quality, etc.), *trip attributes* (such as trip purpose, travel time, etc.), and *circumstances* (such as weather conditions, time of day, traffic information, etc.). However, route choice decisions might not be exclusively dependent on these observable variables, but also on latent variables, which cannot be directly observed, and measured, such as attitudes, perceptions, and lifestyle preferences (Gärling et al., 1998; Hurtubia et al., 2014; McFadden, 1986, 1999). Since every decision maker may have a different perception of these variables, they are considered to be intrinsically subjective (Raveau et al., 2010). The explicit incorporation of these latent constructs in the choice process improves the explanatory power of these models (Ben-Akiva et al., 2002; Prato et al., 2012; Walker, 2001).

Accordingly, different segments of the population, characterized by some of these latent constructs, might also have different choice behaviours (Hurtubia et al., 2014). However, the latent behavioural heterogeneity among the population has mostly been ignored by assuming that all the individuals in the sample population have similar levels of driving experience, spatial knowledge, familiarity with the road network, ability to process information, motivation to compare all the considered alternatives, etc. Ignoring these sources of heterogeneity could reduce the explanatory power of the model and introduce errors in model's forecasts (Ben-Akiva et al., 1993).

Traditional discrete choice models neglect the explicit incorporation of these amorphous constructs and implicitly capture the behavioural aspect of choice process (Kamargianni et al., 2015; Walker, 2001). In order to further improve this underlying aspect of choice modelling and their explanatory power, the explicit incorporation of latent variables such as attitudes and perceptions into the choice model has been proposed (Walker, 2001). Despite the appeal of this framework, its application in route choice studies remains rare. This can be mostly related to the fact that collecting behavioural data is cumbersome and time consuming (Sarkar & Mallikarjuna, 2017). Route choice studies are rarely based on data collected for route choice purposes. Most of the studies using Revealed Preference (RP) data are mostly based on either travel surveys or GPS data, where the presence of behavioural and attitudinal data is scarce. Moreover, studies based on Stated Preference (SP) data mostly focus on observable attributes and avoid attitudinal questions to minimize respondents' burden. Despite all these challenges, why do we still need to estimate such complex models? According to Walker (2001), these models have better prediction abilities, correct for cognitive biases, verify behavioural hypotheses regarding the decision making process, and provide a benchmark to evaluate the performance of more parsimonious models.

In this paper, we examine the route choice behaviour of frequent versus occasional drivers by presenting a comprehensive framework to explicitly incorporate latent behavioural constructs as well as a probabilistic segmentation of the population based on drivers' perceptions and preferences. For this purpose, we evaluate the role of the underlying behavioural constructs on drivers' route choice decisions using the Integrated Choice and Latent Variable (ICLV) framework described by Walker (2001) and adapted to route choice studies by Prato et al. (2012). To properly incorporate the effect of segment heterogeneity and to distinguish between choice behaviours of the different classes of our sample population, we estimate the ICLV model within a Latent Class (LC) framework using a full information estimation approach (Bierlaire, 2016). Moreover, we use ordinal logit models, suggested by Hurtubia et al. (2014), as measurement equations to relate the answers to the psychometric indicators with characteristics of the decision maker in order to improve the characterization of latent classes.

For this purpose, we have designed and implemented a web-based survey to collect behavioural data on drivers' route choices, using behavioural questions on psychometric indicators such as perceptions and attitudes. The Extended Path-Size Logit (EPSL) model, proposed by Frejinger et al. (2009), is adopted to model individuals route choice behaviour, and the Metropolis-Hastings (MH) algorithm presented by Flötteröd and Bierlaire (2013) is used to generate route alternatives for the observed choices.

The rest of this paper is organized as follows. First, we thoroughly present LC and ICLV models and discuss in detail the econometric formulation of the LC-ICLV model. Then, the data collection effort is described, the model specifications are presented and the choice set generation process is discussed. In the end, we highlight the most significant findings of this study, underscore its limitations, and suggest further research directions.

7.2 Econometric Model Formulation

In this section, we first present the general framework of Latent Class (LC) models, followed by a brief illustration of the Integrated Choice and Latent Variable (ICLV) modelling framework. Then, in order to incorporate both the effects of behavioural indicators and taste heterogeneity across population segments, these two models are merged to form the proposed LC-ICLV framework.

7.2.1 Latent Class (LC) discrete choice model

LC choice models assume that different segments of the population exhibit different choice behaviours, due to different lifestyles, preferences and attitudes, which are not directly observed by the analyst (Hurtubia et al., 2014; Yazdizadeh, 2016). It improves the standard multinomial logit model, which considers the same preference structure across the population, by bringing class-specific heterogeneity and considering a probabilistic class membership function which usually depends on individuals' characteristics (Wen & Lai, 2010). Therefore, a class-specific utility function U_{in}^{s} and a class-membership function F_{ns} are defined to calculate the probability of an individual n choosing alternative i conditional on class s, $P_n(i|s)$, and the probability of individual n belonging to class s, $P_n(s)$, respectively. To be consistent with the existing literature, and for the sake of simplicity and clarity, we follow the same notation adopted by Hurtubia et al. (2014) to discuss the LC formulation. The class-specific utility function is given by:

$$U_{in}^{s} = V^{s}(X_{in}, Z_n, \beta^{s}) + \varepsilon_{in}^{s} \tag{7.1}$$

where $V^{s}(X_{in}, Z_n, \beta^{s})$ is the deterministic part of the class-specific utility function composed of attributes of the alternatives X_{in} and individuals' characteristics Z_n; β^{s} is a class-specific vector of parameters to be estimated; and ε_{in}^{s} is an i.i.d Extreme Value distributed random component accounting for unobserved characteristics. Accordingly, $P_n(i|s)$ is calculated as a logit model:

$$P_n(i|s) = \frac{e^{V^{s}(X_{in}, Z_n, \beta^{s})}}{\sum_{j \epsilon C_s} e^{V^{s}(X_{jn}, Z_n, \beta^{s})}} \tag{7.2}$$

where C_s is the set of considered alternatives by individuals in class s. Since individuals cannot be deterministically assigned to latent classes, a class-membership probability is calculated through the definition of a class-membership function F_{ns} as:

$$F_{ns} = f(Z_n, \gamma^s) + \xi_{in} \tag{7.3}$$

where γ^s is a class-specific vector of parameters to be estimated, and ξ_{in} is an i.i.d Extreme Value distributed random component so that $P_n(s)$ is calculated as a logit model:

$$P_n(s) = \frac{e^{f(Z_n, \gamma^s)}}{\sum_{r \epsilon S} e^{f(Z_n, \gamma^s)}} \tag{7.4}$$

where S denotes the set of classes. Scale parameters in equations (7.2) and (7.4) are fixed to 1 for identification purposes (Hurtubia et al., 2014). The complete LC choice model is composed of the two above-mentioned components, such that:

$$P_n(i) = \sum_{s \epsilon S} P_n(i|s) P_n(s) \tag{7.5}$$

7.2.2 Integrated Choice and Latent Variable (ICLV) model

To incorporate the effect of latent variables in the decision-making process, the framework introduced by Walker (2001) and adapted for route choice studies by Prato et al. (2012) is presented in this section. For more details on the presented framework, the reader is referred to (Ben-Akiva et al., 2002; Prato et al., 2012; Walker & Ben-Akiva, 2002; Walker, 2001).

This hybrid modelling framework, also known as the Integrated Choice and Latent Variable (ICLV) model, incorporates psychometric data as indicators of latent variables in the estimation process. Indicators are obtained from drivers responses to behavioural questions of the survey. The model consists of two components: a latent variable model and a choice model. Each component incorporates *structural* as well as *measurement* equations. A *measurement equation* links an unobserved variable to its observable indicators, and a *structural equation* links observable and latent variables to the perceived utility. In the latent variable model, structural equations relate latent variables to explanatory variables, and measurement equations link latent variables to observable indicators. The choice model consists of structural equations relating observable and latent variables to the utility of each route alternative, and measurement equations, which link the unobservable utility to drivers' route choices (Ben-Akiva et al., 2002; Prato et al., 2012). It is worth mentioning that observable indicators and drivers route choices are considered as the manifestation of latent variables and the unobservable utility, in the latent variable and choice models, respectively (Ben-Akiva et al., 2002; Walker, 2001).

In the latent variable component of this hybrid framework, the distribution of the latent variables given the observed variables $f_1(X_n^*|X_n; \lambda, \Sigma_\omega)$ is expressed through the following structural equation:

$$X_n^* = h(X_n; \lambda) + \omega_n \text{ and } \omega_n \backsim D(0, \Sigma_\omega) \tag{7.6}$$

where X_n^* and X_n are vectors of latent and observed variables, respectively; λ is a vector of parameters to be estimated; h is usually a linear in parameter function; ω_n is the random disturbance term with distribution D and a covariance matrix specified by Σ_ω.

The distribution of the utilities in the choice model $f_2(U_n|X_n, X_n^*; \beta, \Sigma_\varepsilon)$ is also specified by a structural equation of the following form:

$$U_n = V(X_n, X_n^*; \beta) + \varepsilon_n \text{ and } \varepsilon_n \backsim D(0, \Sigma_\varepsilon) \tag{7.7}$$

where U_n is the vector of route utilities for alternatives; $V(X_n, X_n^*; \beta)$ is the systematic part of the utility, which is usually a linear in parameter function, dependent on both observable and latent variables; β is a vector of parameters to be estimated; ε_n is the random disturbance term with distribution D and a covariance matrix specified by Σ_ε.

To formulate the distribution of the indicators conditional on the value of latent variables $f_3(I_n|X_n, X_n^*; \alpha, \Sigma_\upsilon)$, measurement equations are expressed as:

$$I_n = m(X_n, X_n^*; \alpha) + \upsilon_n \text{ and } \upsilon_n \backsim D(0, \Sigma_\upsilon) \tag{7.8}$$

where I_n is the vector of indicators; α is a vector of parameters to be estimated; m is usually a linear in parameter function; υ_n is the random disturbance term with distribution D and a covariance matrix specified by Σ_υ. Assuming utility maximization, the measurement equation of the choice model is expressed as a function of the utilities:

$$y_{in} = \begin{cases} 1, & if\ U_{in} \geq U_{jn}\ \forall\ j \neq i \\ 0, & otherwise \end{cases} \tag{7.9}$$

where y_{in} indicates whether route i is chosen by individual n or not. The hybrid model is estimated using maximum likelihood techniques.

Considering the choice probability of selecting a route by taking into account latent variables as $P(y_n|X_n, X_n^*; \beta, \Sigma_\varepsilon)$, and assuming independent error components for structural equations (ω_n, ε_n), the likelihood function is then expressed as the integral of the choice model over the distribution of the latent constructs:

$$P(y_n|X_n; \beta, \Sigma_\omega, \Sigma_\varepsilon) = \int_{X^*} P(y_n|X_n, X^*; \beta, \Sigma_\varepsilon)\, f_1(X^*|X_n; \lambda, \Sigma_\omega) dX^* \tag{7.10}$$

To improve the accuracy of the estimates, psychometric data is used alongside the observed preference attributes as indicators of latent psychological factors I_n (Walker & Ben-Akiva, 2002).Therefore, to achieve the final likelihood equation of the hybrid model, the density function of indicators $f_3(I_n|X_n, X_n^*; \alpha, \Sigma_\upsilon)$, should be incorporated into the joint probability function.

$$\begin{aligned} &P(y_n, I_n|X_n; \alpha, \beta, \lambda, \Sigma_\varepsilon, \Sigma_\upsilon, \Sigma_\omega) \\ &\qquad = \int_{X^*} P(y_n|X_n, X^*; \beta, \Sigma_\varepsilon)\, f_3(I_n|X_n, X^*; \alpha, \Sigma_\upsilon) f_1(X^*|X_n; \lambda, \Sigma_\omega) dX^* \end{aligned} \tag{7.11}$$

To have a logit form function, we assume that the error terms ε_n are independently and identically Gumbel distributed. However, error terms ω_n and υ_n are assumed to be normally and independently distributed yielding orthogonal latent variables. Density functions of latent variables and indicators are given as

$$f_1(X_n^*|X_n; \lambda, \sigma_\omega) = \prod_{l=1}^{L} \frac{1}{\sigma_{\omega_l}} \Phi\left(\frac{X_{ln}^* - h(X_n; \lambda_l)}{\sigma_{\omega_l}}\right) \text{ and} \tag{7.12}$$

$$f_3(I_n|X_n, X_n^*; \alpha, \sigma_\upsilon) = \prod_{r=1}^{R} \frac{1}{\sigma_{\upsilon_r}} \Phi\left(\frac{I_{rn} - m(X_n, X_n^*; \alpha_r)}{\sigma_{\upsilon_r}}\right) \tag{7.13}$$

where Φ is the standard normal density function; σ_ω and σ_υ are standard deviations of the error terms ω_n and υ_n, respectively; R is the number of indicators; and, L is the number of latent variables.

7.2.3 LC_ICLV choice model

In order to incorporate both the effects of segment heterogeneity and latent variables in the decision-making process, we considered an ICLV model as the choice component, $P_n(i|s)$, of the LC model. According to Hurtubia et al. (2014) and Bierlaire (2016), since psychometric indicators are collected using a five levels Likert scale, measurement equations should be represented by an ordered discrete variable H taking values 1 to 5:

$$H = \begin{cases} 1, & \text{if } I_n < \tau_1 \\ 2, & \text{if } \tau_1 \ \leq I_n < \tau_2 \\ 3, & \text{if } \tau_2 \ \leq I_n < \tau_3 \\ 4, & \text{if } \tau_3 \ \leq I_n < \tau_4 \\ 5, & \text{if } \tau_4 \ \leq I_n \end{cases} \tag{7.14}$$

where I_n is defined by Equation 7.8, and $\tau_1, \ldots, \tau_4$ are parameters to be estimated, such that:

$$\tau_1 \leq \tau_2 \leq \tau_3 \leq \tau_4 \tag{7.15}$$

Considering a normal distribution for the error term in I_n, υ, the probability of a given response is calculated by an ordered probit model:

$$\Pr(I_i = \ j_i) = \Pr(\tau_{i-1} \leq I_i \leq \tau_i) = \ \mathcal{F}_\upsilon(\tau_i) - \ \mathcal{F}_\upsilon(\tau_{i-1}) \tag{7.16}$$

where $j_i \in \ \{1, 2,3,4,5\}$, and $\mathcal{F}_\upsilon$ is the cumulative distribution function of the error term υ.

To achieve consistent and efficient estimates, a full information estimation method has been employed to simultaneously estimate the parameters of the choice model, class-membership, and responses to psychometric indicators (Ben-Akiva et al., 2002; Bierlaire, 2016; Cantillo, Arellana, & Rolong, 2015; Walker, 2001).

$$P_n(i) = \sum_{s \epsilon S} \left\{ \left(\int_{X^*} P(y_n|X_n, X^*; \beta, \Sigma_\varepsilon) \prod_k P_n(I_k|X_n, X^*; \alpha, \Sigma_\upsilon) \, f_1(X^*|X_n; \lambda, \Sigma_\omega) dX^* \right) P_n(s) \right\} \tag{7.17}$$

7.3 Empirical Application

The proposed LC-ICLV framework is used to test the hypothesis of difference between frequent versus occasional drivers' route choice behaviours. For this purpose, we have used revealed preference data containing behavioural indicators, collected via a web-based survey. This section presents the data, the survey participants, and the model specification.

7.3.1 Data

Data has been collected through a revealed preference web-based survey, conducted in Montreal, Quebec in 2017, which targets drivers residing and driving in the Greater Montreal Area. The survey has been designed to identify behavioural and attitudinal factors affecting drivers' route choice behaviours.

It first collects typical information on sociodemographic and socioeconomic characteristics of participants such as age, gender, educational attainment, type of work, salary, household size, and number of cars in the household. Then, respondents are asked to specify on a geographical map, the destination point to which they drive most frequently and all the alternative routes which they consider for the specified trip. Moreover, they are asked to specify the frequency of the trip on a weekly basis. Additionally, they are requested to provide their level of agreement to a list of statements designed to reveal drivers' attitudes, preferences and perceptions towards choosing a route (Atasoy et al., 2013; Prato et al., 2012; Walker, 2001). Responses are provided on a five point Likert scale ranging from total agreement to total disagreement. In total, 74 questions were asked and considering a 95th percentile threshold, the average response time was found to be around 16 minutes. A thorough description of the data collection effort, recruitment methods, participants' characteristics, response rates, and dropouts are described in (Alizadeh, Bourbonnais, Morency, Farooq, & Saunier, 2017).

In this study, only respondents with a trip origin and destination inside the Island of Montreal and declaring a trip frequencies of one time per week (137 observations) or more than five times per week (88 observations) have been selected for estimation, adding up to a total of 225 observations. Table 7.1 provides comprehensive details on respondents' sociodemographic and socioeconomic characteristics.

Table 7.1: Characteristics of the Survey Participants

Variable categories	N	%	Variable categories	N	%
Gender			*Household size*		
Male	118	52.4	1	54	24.0
Female	107	47.6	2	81	36.0
Age (years old)			3	41	18.3
Young (15 to 39)	120	53.3	4	37	16.4
Middle age (40 to 59)	85	37.7	+5	12	5.3
Old (more than 60)	20	9.0			
Occupation			*Income (Thousand CAD per capita)*		
Full time worker	169	75.1	Low (Less than 35)	58	25.8
Partial time worker	18	8.0	Medium (35 to 75)	88	39.1
Student	22	9.8	High (more than 75)	35	15.6
Retired	9	4	Not declared	44	19.6
House-wife/husband	2	0.9	*Household car number*		
Other	5	2.2	0	57	25.3
Education			1	129	57.3
None	0	0	2	32	14.2
Less than university	26	11.6	3	6	2.7
University	197	87.6	4	0	0.0
Other	2	0.8	5	1	0.4

7.3.2 Model Specification

The choice model incorporated into the likelihood function is an Extended Path-Size Logit (EPSL) model, proposed by Frejinger et al. (2009). This model uses a correction factor in the deterministic part of the utility function to take into account the correlation of each alternative with all the possible paths in the true choice set. The conditional probability takes the following form:

$$P_{EPSL}(i|C_n) = \frac{e^{(V_{in}+lnEPS_{in})+ln(\frac{k_{in}}{q(i)})}}{\sum_{j\epsilon C_n} e^{(V_{jn}+lnEPS_{jn})+ln(\frac{k_{jn}}{q(j)})}} \tag{7.18}$$

where $P_{EPSL}(i|C_n)$ is the conditional probability of user n choosing alternative i from the choice set C_n, μ is a scale parameter, $ln(\frac{k_{in}}{q(i)})$ is the sampling correction factor, $q(i)$ is the sampling probability of path i, and k_{in} is the empirical frequency or the actual number of times path i is drawn. It is also worth mentioning that the utility is a function of both observable and latent variables $V_{in} = V_{in}(X_n, X_n^*; \beta)$. The EPS factor is defined as:

$$EPS_{in} = \sum_{a \in \Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j \epsilon \varphi_n} \delta_{aj} \omega_{jn}} \tag{7.19}$$

where L_a and L_i denotes the length of link a and path i, Γ_i represents the set of road segments in path i, φ_n denotes the considered choice set, and δ_{aj} is the link-path incident binary variable, which is 1 if link a is on path i, and 0 otherwise(i.e., $\sum_{j \epsilon \varphi_n} \delta_{aj}$ indicates the total number of route alternatives in the choice set sharing link a. ω_{jn} is an extension factor with a value equal to 1 if $\delta_{aj} = 1$ or $q(j)B_n \geq 1$, and $1/(q(j)B_n\ )$ otherwise; where B_n denotes the total number of paths drawn with replacement from the universal choice set. Measurement equation for indicator k is formulated as given by:

$$I_k = \ \gamma_k + \ \alpha_k X^* + \upsilon_k \text{ and } \upsilon_n \backsim N(0, \Sigma_{\upsilon_k}) \tag{7.20}$$

where I_k represents the k^{th} psychometric indicators, X^* denotes the latent variable, γ_k and α_k are parameters to be estimated, and υ_k is a normally distributed random term. The structural equation of the latent variable component m (LV_m) of the model is defined to associate latent variables to individual characteristics. Since the main purpose of this study is to compare frequent versus occasional drivers' route choice behaviours, the following form is defined:

$$LV_{mn} = \ \lambda_{m1} FREQ_n + \ \lambda_{m2} CAR_n + \omega_{mn} \text{ and } \omega_{mn} \backsim N(0, \Sigma_\omega) \tag{7.21}$$

where $FREQ_n$ is equal to 1 if the respondent drives more than 5 times per week to the specified destination, and equal to 0 when he/she makes the trip once per week, CAR_n specifies the number of car per household, ω_{mn} is a variable specific normally distributed random term, $and\ \ \lambda_{m1}$ and λ_{m2} are parameters to be estimated.

Class specific utility functions (structural equations) of the choice model, are specified to associate route attributes and behavioural traits to the utility perceived by individual n ($C1$ and $C2$ represent Class 1 and Class 2, respectively). According to Prato et al. (2012), the effect of latent variables on drivers route choice can be captured through an interaction term between latent variables and route attributes. Accordingly, the systematic part of the choice model's structural equations for each class is empirically defined by:

$$V_n^{C1} = \beta_{ROUTE\_LEN}^{C1} \times CON_n ROUTE\_LEN_n + \beta_{HGW\_PERC}^{C1} \times CON_n HGW\_PERC_n + \beta_{TURN\_LEN}^{C1} \times CON_n TURN\_LEN_n + \beta_{REAL\_TT}^{C1} \times CON_n REAL\_TT_n \tag{7.22}$$

$$V_n^{C2} = \beta_{ROUTE\_LEN}^{C2} \times CAU_n ROUTE\_LEN_n + \beta_{HGW\_LEN}^{C2} \times CAU_n HGW\_LEN_n + \beta_{TURN\_LEN}^{C2} \times CAU_n TURN\_LEN_n + \beta_{REAL\_TT}^{C2} \times CAU_n REAL\_TT_n \tag{7.23}$$

where $ROUTE\_LEN$ is the route length in meters, HGW_PERC signifies the highway portion of the route in percentage, $TURN\_LEN$ denotes the number of turns per kilometer, and $REAL\_TT$ indicates the estimated real travel time in minutes. CON_n and CAU_n are the two latent variables considered in this study and defined in the next section through Equation 7.21. To distinguish between the two different classes, the class-membership functions are defined to be a function of individuals' socioeconomic characteristics:

$$f(Z_n, \gamma^1) = \gamma_{INC\_H}^1 INC\_H_n + \gamma_{FAM\_EXP}^1 FAM\_EXP_n + \gamma_{AGE\_MIDDLE}^1 AGE\_MIDDLE_n + \gamma_{LIVE\_HOME}^1 LIVE\_HOME_n \tag{7.24}$$

$$f(Z_n, \gamma^2) = 0 \tag{7.25}$$

INC_H is set to 1 where the income is high and 0 otherwise, FAM_EXP denotes the level of familiarity with road network of Montreal and the level of individuals' driving experience, AGE_MIDDLE is 1 where the participant is middle-aged and 0 otherwise, $LIVE\_HOME$ specifies the duration (in years) that the participant has been living in the same address.

7.3.3 Choice set generation

Since it is not possible to enumerate all the possible paths between an OD pair in a real-world road network, implicit or explicit path generation techniques are usually used to simulate drivers' considered set of route alternatives. We adopt the Metropolis-Hastings (MH) based route generation algorithm, proposed by Flötteröd and Bierlaire (2013), which uses an underlying Markov Chain process to sample alternatives. Its major advantage over conventional methods (e.g. link labelling, link elimination, etc.) is that it provides researchers with path sampling probabilities, so that model estimates based on these sets are not biased. In this study, 19 choice alternatives were generated for each observation and the chosen alternative has been added to the choice set, where it was not generated by the MH algorithm (Dhakar & Srinivasan, 2014; Elgar et al., 2009; Frejinger et al., 2009; Habib et al., 2013; Hess et al., 2015; McFadden, 1978; Prato et al., 2012).

7.4 Results and discussion

A preliminary Principal Factor Analysis (PFA) on psychometric indicators demonstrated the distinction of the two prominent components. The first component corresponds to drivers with better memory and sense of direction, willing to minimize their travel time by driving on highways, avoiding traffic lights, construction sites and congestion, and more open to change their habitual route and try new ones. We refer to this factor as the *Consciousness* (CON) attitude. In the second component, entitled the *Cautiousness* (CAU) attitude, drivers are more inclined towards local and scenic routes, and are more prone to avoid trucks and narrow lanes as much as possible. Table 7.2 presents these latent attitudes, indicators associated to them and their respective factor loadings proportions.

Table 7.2: Description of Indicators and Factor Loading Coefficients

ID	Indicator	Description	Factor loadings	
			Consciousness	Cautiousness
1	NRW_LN	Avoiding narrow lanes		0.443996
2	AV_TRK	Avoiding routes with high number of trucks		0.508139
3	SCEN	Preferring scenic routes		0.644226
4	LOCAL	More willing to take local routes		0.749473
5	HGW	More willing to take highways	0.429807	-0.590120
6	PAV	Importance of pavement quality	0.477136	
7	SHTC	Searching for route shortcuts	0.617959	
8	TLIGHT	Avoiding traffic lights	0.677496	
9	TT_MIN	Minimizing travel time	0.428299	
10	DRV_STK	Preferring longer routes rather than being stuck in traffic jams on shorter ones	0.460730	
11	NEW_RT	Willingness to try new routes	0.482361	
12	AV_CONST	Avoiding construction sites	0.418859	
13	SOD	Having a good sense of direction	0.401937	
14	CHANGE_RT	Changing route in case of accidents	0.562993	
15	MEM	Having a good memory	0.482990	
16	HG_SPD	Preferring higher speed limits	0.481988	
17	DIR	Preferring more direct routes	0.341694	
18	TT_REL	Having more reliable travel time	0.441834	
19	HABIT	Taking the same route repeatedly	-0.341700	

For the *Cautiousness* attitude, indicators 1, 2, 3, and 4 were found to significantly improve the LC-ICLV model, while for the *Consciousness* attitude, all the indicators except 6, 11, 16 and 17 were included in the model estimation.

The BIOGEME software package (Bierlaire, 2003; Bierlaire & Fetiarison, 2009) has been used to jointly estimate the parameters of the measurement equations, relating indicators to the latent components, the class-membership functions and the choice model, using a full estimation approach described in detail in (Bierlaire, 2016). An incremental exploratory estimation approach has been used to obtain good initial values from simpler model specifications and reduce the estimation time of more complex ones (Hurtubia et al., 2014).

Estimation results are presented in Table 7.3 for measurement equations of the latent variable component of the model. One of the parameters in each latent variable is constrained to 1 for identification purposes (Ben-Akiva et al., 2002; Walker, 2001). As expected from the preliminary factor analysis results, the latent variable *Consciousness* is positively correlated to the tendency of minimizing the travel time, taking shortcuts and highways, avoiding traffic lights, changing route in case of accidents and trying new ones, avoiding lights, taking routes with more reliable travel times, and having a good memory and sense of direction. Intuitively, it is negatively linked to the habit of taking the same route. The correlation of the latent variable *Cautiousness* is positively linked to the willingness of avoiding trucks and narrow lanes and the inclination towards local routes and scenic routes.

Table 7.3: Estimates of the Measurement Equations of the ICLV Model

Attribute	Parameter	Estimate	t-Stat	Attribute	Parameter	Estimate	t-Stat
Consciousness				*Cautiousness*			
CHANGE_RT	α_{CHANGE_RT}	1	-	LOCAL	α_{LOCAL}	1	-
	σ_{CHANGE_RT}	1	-		σ_{LOCAL}	1	-
AV_CONST	α_{AV_CONST}	0.202	2.15	AV_TRK	α_{AV_TRK}	0.634	1.76*
	σ_{AV_CONST}	1.09	15.14		σ_{AV_TRK}	1.23	14.10
DRV_STK	α_{DRV_STK}	0.929	9.77	NRW_LN	α_{NRW_LN}	1.71	2.96
	σ_{DRV_STK}	1.10	13.88		σ_{NRW_LN}	1.11	13.16
HGW	α_{HGW}	0.844	7.64	SCEN	α_{SCEN}	0.862	3.04
	σ_{HGW}	1.24	15.59		σ_{SCEN}	1.08	14.71
HABIT	α_{HABIT}	-0.645	-6.87				
	σ_{HABIT}	1.05	14.63				
TRF_LGH	α_{TRF_LGH}	0.103	1.39*				
	σ_{TRF_LGH}	0.906	13.98				
MEM	α_{MEM}	1.07	10.23				
	σ_{MEM}	1.12	12.67				
SOD	α_{SOD}	1.02	7.53				
	σ_{SOD}	1.17	13.19				
SHTC	α_{SHTC}	0.584	6.65				
	σ_{SHTC}	1.02	13.86				
TT_REL	α_{TT_REL}	0.420	5.56				
	σ_{TT_REL}	0.897	16.40				
TT_MIN	α_{TT_MIN}	1.28	9.47				
	σ_{TT_MIN}	1.31	13.18				

* Not statistically significant with $p < 0.05$

Table 7.4 illustrates the estimates for the structural equations of the latent variable component of the model. Results illustrate that being a frequent driver (FREQ) and living in highly motorized households (CAR) are associated with the *Consciousness* type of attitude, and occasional, i.e. less frequent, drivers having fewer cars per household seem expectedly to be more related with the Cautiousness type of attitude.

Table 7.4: Estimates of the Structural Equations of the ICLV Model

Factors	Latent variables			
	Consciousness		Cautiousness	
	Estimate	t-Stat.	Estimate	t-Stat.
FREQ	0.487	2.99	-0.605	-2.06
CAR	1.940	10.35	-0.185	-5.58

Estimation results for the choice model and the class-membership function is illustrated in Table 7.5. A latent class model without the ICLV component has also been estimated as a benchmark to compare the results. The same specification of the utility function and class-membership equations have been used in both models. It can be concluded that the inclusion of behavioural traits in the LC model significantly improves its fit over the data.

Table 7.5: Estimates of the Choice Model and Class-Membership Functions

Variables	LC-ICLV Estimate	LC-ICLV t-Stat.	LC Estimate	LC t-Stat.
Choice Model				
$\beta^{C1}_{ROUTE\_LEN}$	-0.003	-6.57	-0.004	-2.13
$\beta^{C2}_{ROUTE\_LEN}$	-0.001	-2.09	-0.001	-2.44
$\beta^{C1}_{HGW\_PERC}$	0.207	5.81	0.338	2.44
$\beta^{C2}_{HGW\_PERC}$	-0.222	-2.83	-0.044	-2.14
$\beta^{C1}_{TURN\_LEN}$	-1.130	-3.5	-3.530	-5.02
$\beta^{C2}_{TURN\_LEN}$	-0.577	-6.33	-0.290	-1.71*
$\beta^{C1}_{REAL\_TT}$	-0.168	-1.91*	-0.256	-1.09*
$\beta^{C2}_{REAL\_TT}$	0.089	1.47*	0.984	0.88*
β^{C1}_{EPS}	1.50	8.32	0.953	2.94
β^{C2}_{EPS}	1.10	18.95	0.793	12.34
Latent Class				
$\gamma^{1}_{INC\_H}$	1.24	2.05	1.44	2.29
$\gamma^{1}_{FAM\_EXP}$	0.0148	3.82	0.0371	2.46
$\gamma^{1}_{AGE\_MIDDLE}$	0.698	1.59	0.879	2.15
$\gamma^{1}_{LIVE\_HOME}$	0.012	3.34	0.0638	2.14
Number of parameters	44		14	
Initial log-likelihood	-9512.9		-2716.7	
Final log-likelihood	-4139.73		-1409.9	
Rho-square	0.564		0.481	

* Not statistically significant with $p < 0.05$

In the class-membership model, it can be noticed that the probability of belonging to Class 1 increases with being a middle-age individual (AGE_MIDDLE), having high income (INC_H), higher duration of living in the same home ($LIVE\_HOME$), and being more experienced in driving and more familiar with Montreal's road network (FAM_EXP). Individuals in this class are mostly affected by the *Consciousness* type of attitude, and show a positive relationship with driving frequency and number of cars per household (see Table 7.4). Hence, individuals in Class 1 are entitled as "Frequent Drivers" and accordingly, those in Class 2 are considered as "Occasional Drivers".

A closer look at the parameters of the utility function reveals that travel distance, number of turns, and travel time have their expected signs and negatively affect the perceived utility (except for the travel time parameter of Class 2). Intuitively, drivers prefer shorter routes, with lower travel times and fewer number of turns.

Due to the latent segmentation of the population, different sensitivities are noticed regarding these attributes. Logically, frequent drivers are more sensitive to travel distance, number of turns, and travel time than occasional drivers. The positive sign for the travel time parameter of Class 2 further implies that occasional drivers are less sensitive to travel time and may take their habitual route or a more scenic route, even if the travel time is longer. Moreover, the sign for highway percentage implies that frequent drivers are more willing to use highways, while occasional drivers tend to avoid them by taking local routes. This behaviour is consistent with the level of agreement answers given to questions related to freeway and local route usages. In other words, drivers associated with the consciousness attitude gave more agreement to the question related to freeway usage, while drivers related to the cautiousness attitude agreed more with the statement of driving on local routes. Moreover, the positive signs of β_{EPS}^{C1} and β_{EPS}^{C2} is a negative correction of the utility of overlapping routes, giving a higher chance to less similar alternatives to be chosen.

7.5 Summary and Conclusions

Route choice modelling has an indispensable role in transportation planning and simulation. It provides insights on drivers' perceptions of route characteristics and prepare the ground for the simulation and forecast of traveller's behaviour under hypothetical scenarios. Incorporation of latent behavioural traits and segment taste heterogeneity improves the behavioural aspect of the modelling process, and consequently increase model's predictions abilities (Walker & Ben-Akiva, 2002; Walker, 2001).

Discrete choice models have been extensively used over the past few decades to study route choices and factors affecting them. These factors can be classified into two broad categories: observable and latent. Observable factors are those that are tangible and can be directly observed, such as alternatives' features and drivers' characteristics. However, it has been well established that latent variables such as attitudes, perceptions and lifestyle preferences play a major role in the decision making process (Gärling et al., 1998; Kamargianni et al., 2015; McFadden, 1999; Muñoz et al., 2016; Sarkar & Mallikarjuna, 2017). For instance, in addition to observable factors such as travel time, travel distance, and number of turns, route choices might also depend on driver's experience, familiarity with the road network, and safety concerns.

In this paper, we examine the route choice behaviour of frequent versus occasional drivers by presenting a comprehensive framework to explicitly incorporate latent behavioural constructs as well as a probabilistic segmentation of the population based on drivers' perceptions and preferences. We introduced an ICLV model with an EPSL choice component into a LC model and adopted ordinal logit models to express measurements equations of the latent variable component of the ICLV model. A MH based route generation method is used to generate the considered choice set of each observation (Flötteröd & Bierlaire, 2013).

Data has been collected through a revealed preference web-based survey, conducted in Montreal, Canada in 2017. The survey has been designed to identify behavioural and attitudinal factors affecting drivers' route choice behaviours. The modelling dataset includes 225 drivers residing and driving in the Greater Montreal Area.

Two major behavioural traits have been observed among the drivers, namely *Consciousness* and *Cautiousness* factors. Drivers related to the *Consciousness* factor tend to minimize their travel time, try shortcuts and highways, avoid traffic lights, change route in case of accidents and try new ones, avoid lights, take routes with more reliable travel times, and have a good memory and sense of direction. However, the latent factor *Cautiousness* is positively linked to the willingness of avoiding trucks and narrow lanes and the inclination towards local and scenic routes.

Expectedly, the choice model illustrated that in general, drivers prefer shorter routes, with lower travel times and fewer number of turns. However, the comparison between frequent and occasional drivers showed that frequent drivers are more sensitive to travel distance, number of turns, and travel time than occasional drivers. Moreover, it showed that occasional drivers are not necessarily travel time minimizers and prefer to avoid driving on highways. The different signs for the travel time parameter and the highway percentage parameter across the two segments further imply the necessity of incorporating the effect of segment heterogeneity and to distinguish between choice behaviours of the different segment of the population. Finally, results demonstrated that the inclusion of behavioural traits in the LC model significantly improves its fit over the data.

In order to improve the findings of the presented work, a further research direction could be the inclusion of different attitudinal indicators, and more complex specifications of the class-membership functions in order to better characterize the compared population segments. A complementary research could be to evaluate the prediction power of the proposed model and compare it to simpler models such as the LC model. Another interesting area to explore would be the extension of the Recursive Logit model (Fosgerau et al., 2013; Zimmermann, Mai, & Frejinger, 2017b) by including taste heterogeneity in terms of latent variables and latent classes. Finally, the proposed framework could be tested on other datasets to compare the choice behaviour of different segments of the population.

Acknowledgement

The authors would like to express their gratitude to Pierre-Léo Bourbonnais for the design of the survey framework as well as the partners of the Mobilité Research Chair of Polytechnique Montréal for their financial support, for providing access to data for research purposes and their help in the dissemination of the survey.

References

Alizadeh, H., Bourbonnais, P.L., Morency, C., Farooq, B., Saunier, N., 2017a. An Online Survey to Enhance the Understanding of Car Drivers Route Choices, International Conference on Transport Survey Methods, Montreal, Canada.

Alizadeh, H., Farooq, B., Morency, C., Saunier, N., 2017b. On the role of bridges as anchor points in route choice modeling. Transportation, 1-26.

Atasoy, B., Glerum, A., Bierlaire, M., 2013. Attitudes towards mode choice in Switzerland. disP - The Planning Review 49(2), 101-117.

Ben-Akiva, M., Bolduc, D., Bradley, M., 1993. Estimation of travel choice models with randomly distributed values of time.

Ben-Akiva, M., Walker, J., Bernardino, A.T., Gopinath, D.A., Morikawa, T., Polydoropoulou, A., 2002. Integration of choice and latent variable models. Perpetual motion: Travel behaviour research opportunities and application challenges, 431-470.

Bierlaire, M., 2003. BIOGEME: a free package for the estimation of discrete choice models, Swiss Transport Research Conference.

Bierlaire, M., 2016. Estimating choice models with latent variables with PythonBiogeme, EPFL.

Bierlaire, M., Fetiarison, M., 2009. Estimation of discrete choice models: extending BIOGEME, Swiss Transport Research Conference (STRC).

Bierlaire, M., Frejinger, E., 2008. Route choice modeling with network-free data. Transportation Research Part C: Emerging Technologies 16(2), 187-198.

Cantillo, V., Arellana, J., Rolong, M., 2015. Modelling pedestrian crossing behaviour in urban roads: A latent variable approach. Transportation Research Part F: Traffic Psychology and Behaviour 32, 56-67.

Dalumpines, R., Scott, D.M., 2017. Determinants of route choice behavior: A comparison of shop versus work trips using the Potential Path Area-Gateway (PPAG) algorithm and Path-Size Logit. Journal of Transport Geography 59, 59-68.

Dhakar, N., Srinivasan, S., 2014. Route choice modeling using GPS-based travel surveys. Transportation Research Record: Journal of the Transportation Research Board(2413), 65-73.

Dhakar, N.S., 2012. Route Choice Modeling Using GPS Data. University of Florida.

Dougherty, M., 1995. A review of neural networks applied to transport. Transportation Research Part C: Emerging Technologies 3(4), 247-260.

Elgar, I., Farooq, B., Miller, E., 2009. Modeling location decisions of office firms: Introducing anchor points and constructing choice sets in the model system. Transportation Research Record: Journal of the Transportation Research Board(2133), 56-63.

Flötteröd, G., Bierlaire, M., 2013. Metropolis–Hastings sampling of paths. Transportation Research Part B: Methodological 48, 53-66.

Fosgerau, M., Frejinger, E., Karlstrom, A., 2013. A link based network route choice model with unrestricted choice set. Transportation Research Part B: Methodological 56(Supplement C), 70-80.

Frejinger, E., Bierlaire, M., Ben-Akiva, M., 2009. Sampling of alternatives for route choice modeling. Transportation Research Part B: Methodological 43(10), 984-994.

Gao, S., Frejinger, E., Ben-Akiva, M., 2010. Adaptive route choices in risky traffic networks: A prospect theory approach. Transportation Research Part C: Emerging Technologies 18(5), 727-740.

Gärling, T., Gillholm, R., Gärling, A., 1998. Reintroducing attitude theory in travel behavior research: The validity of an interactive interview procedure to predict car use. Transportation 25(2), 129-146.

Habib, K.N., Morency, C., Trépanier, M., Salem, S., 2013. Application of an independent availability logit model (IAL) for route choice modelling: Considering bridge choice as a key determinant of selected routes for commuting in Montreal. Journal of Choice Modelling 9, 14-26.

Hess, S., Quddus, M., Rieser-Schüssler, N., Daly, A., 2015. Developing advanced route choice models for heavy goods vehicles using GPS data. Transportation Research Part E: Logistics and Transportation Review 77, 29-44.

Hurtubia, R., Nguyen, M.H., Glerum, A., Bierlaire, M., 2014. Integrating psychometric indicators in latent class choice models. Transportation Research Part A: Policy and Practice 64, 135-146.

Jan, O., Horowitz, A., Peng, Z.-R., 2000. Using global positioning system data to understand variations in path choice. Transportation Research Record: Journal of the Transportation Research Board(1725), 37-44.

Kamargianni, M., Dubey, S., Polydoropoulou, A., Bhat, C., 2015. Investigating the subjective and objective factors influencing teenagers' school travel mode choice – An integrated choice and latent variable model. Transportation Research Part A: Policy and Practice 78, 473-488.

Li, D., Miwa, T., Morikawa, T., Liu, P., 2016. Incorporating observed and unobserved heterogeneity in route choice analysis with sampled choice sets. Transportation Research Part C: Emerging Technologies 67(Supplement C), 31-46.

Manski, C.F., 1977. The structure of random utility models. Theory and decision 8(3), 229-254.

McFadden, D., 1978. Modelling the choice of residential location. Institute of Transportation Studies, University of California.

McFadden, D., 1986. The Choice Theory Approach to Market Research. Marketing Science 5(4), 275-297.

McFadden, D., 1999. Rationality for Economists? Journal of Risk and Uncertainty 19(1), 73-105.

Muñoz, B., Monzon, A., López, E., 2016. Transition to a cyclable city: Latent variables affecting bicycle commuting. Transportation Research Part A: Policy and Practice 84, 4-17.

Prato, C.G., 2009. Route choice modeling: past, present and future research directions. Journal of Choice Modelling 2(1), 65-100.

Prato, C.G., Bekhor, S., Pronello, C., 2012. Latent variables and route choice behavior. Transportation 39(2), 299-319.

Raveau, S., Álvarez-Daziano, R., Yáñez, M., Bolduc, D., de Dios Ortúzar, J., 2010. Sequential and simultaneous estimation of hybrid discrete choice models: some new findings. Transportation Research Record: Journal of the Transportation Research Board(2156), 131-139.

Ridwan, M., 2004. Fuzzy preference based traffic assignment problem. Transportation Research Part C: Emerging Technologies 12(3), 209-233.

Sarkar, P.P., Mallikarjuna, C., 2017. Effect of perception and attitudinal variables on mode choice behavior: A case study of Indian city, Agartala. Travel Behaviour and Society.

Walker, J., Ben-Akiva, M., 2002. Generalized random utility model. Mathematical Social Sciences 43(3), 303-343.

Walker, J.L., 2001. Extended discrete choice models: integrated framework, flexible error structures, and latent variables. Massachusetts Institute of Technology.

Wen, C.-H., Lai, S.-C., 2010. Latent class models of international air carrier choice. Transportation Research Part E: Logistics and Transportation Review 46(2), 211-221.

Xu, H., Zhou, J., Xu, W., 2011. A decision-making rule for modeling travelers' route choice behavior based on cumulative prospect theory. Transportation Research Part C: Emerging Technologies 19(2), 218-228.

Yazdizadeh, A., 2016. A Generic Form for Capturing Unobserved Heterogeneity in Discrete Choice Modelling: Application to Neighborhood Location Choice. Concordia University.

Zimmermann, M., Mai, T., Frejinger, E., 2017. Bike route choice modeling using GPS data without choice sets of paths. Transportation Research Part C: Emerging Technologies 75(Supplement C), 183-196.

CHAPTER 8 ARTICLE 5: FACTORS AFFECTING DRIVERS' CONSIDERATION SET OF ROUTE ALTERNATIVES

Hamzeh Alizadeh[1*], Catherine Morency[1], Nicolas Saunier[1], Bilal Farooq[2]

Paper submitted for publication in *Transportation Research Board*
Submission Date: 15 / February / 2018

[1] Department of Civil, Geological and Mining Engineering, Polytechnique Montréal, C.P. 6079, succ. Centre-Ville Montréal, Canada

[2] Department of Civil Engineering, Ryerson University, 350 Victoria Street, Toronto, Canada

[*] Corresponding author

Abstract

In a two-stage choice modelling process, determining drivers' consideration set of route alternatives is an essential component. The misspecification of the size and composition of the considered choice set may lead to biased estimates and fallacious predicted demand levels. However, the way individual drivers derive their actual consideration set of route alternatives, and factors affecting the size and composition of these choice sets is still unknown. The key role of choice set definition and the need to get some insights regarding drivers' actual consideration set has motivated this research.

Data has been collected through a web-based data collection framework, designed to collect information on drivers' actual consideration set of route alternatives. The survey was conducted in the Greater Montreal Area and information on 506 drivers declaring 988 route alternatives was collected. Then, the effect of six broad categories of factors on the size of drivers' consideration sets has been studied. These categories include personal attributes, declared factors, behavioural indicators, incentives, awareness determinants, and spatial, temporal and environmental components.

Accordingly, four types of behaviour were distinguished with regard to the size of drivers' consideration sets, namely "*Determined cautious drivers*", *"Biased habitual drivers"*, *"Middling impartial drivers"*, and *"Swayable conscious drivers"*. Important factors and behavioural traits

affecting drivers' choice selection behaviour can be embedded in the non-compensatory choice set formation step to generate and select more behaviourally realistic route alternatives. Moreover, the observed choice set can improve model's estimation and prediction efficiency by providing detailed information regarding travellers' preferences.

Keywords: Consideration choice set, Route choice, Revealed preference data, Web-based survey, Choice set generation

8.1 Introduction

Modelling route choice behaviour is one of the most intricate and challenging tasks in transportation studies. The complexity arises from two main challenges, namely the large number of possible routes connecting a given origin destination (OD) pair, and the complex correlation structure among the overlapping routes. Accordingly, discrete choice analysis framework and a two-stage behavioural process is usually adopted to simulate drivers' route choices. This framework assumes that choice set formation and choice from the considered set of alternatives are two distinct mental processes, in which the former precedes the latter (Ben-Akiva & Boccara, 1995; Bovy, 2009). Assuming a two-stage process for modelling route choice decisions, this paper contributes to the line of work addressing the first stage, i.e. the analysis of the considered set of route alternatives.

Choice sets are specified as collections of feasible travel alternatives considered by individual travellers. Defining them is a crucial step in modelling and analysing travel choice behaviour. It has been well-established in the literature that restricting all possible choices (i.e. the universal choice set) to alternatives considered by the decision maker improves the prediction ability of the choice model and represents the choice process in a more realistic way (Horowitz & Louviere, 1995). Previous studies illustrate that parameter estimates change with various definitions of the alternative set. They have also argued that the size of the choice set and its variability greatly affect models' estimates and choice probabilities over the set of alternatives (Bliemer & Bovy, 2008; Geda, 2014; Peters et al., 1995; Prato & Bekhor, 2006, 2007b; Schuessler & Axhausen, 2009; Swait & Ben-Akiva, 1987a).

In route choice studies and real world road networks, it is not feasible to identify the *universal choice set,* i.e. to enumerate all the paths connecting a given OD pair. Hence, deterministic and stochastic route generation techniques have been adopted to create a subset called *master set*, which approximates all the routes that are supposed to be known to the decision maker. However, this set may still be very large and may contain non-feasible and unattractive alternatives. Moreover, drivers' are unrealistically assumed to know all of them and tirelessly compare their attributes to choose the best one. Due to the limited information processing abilities of drivers, spatial and temporal restrictions, and latent traits, such as attitudes and preferences, a set of spatiotemporal constraints and screening rules are adopted to delimit the *consideration set*, which is supposed to represent the actual set of routes from which drivers pick their final choice (Ben-Akiva & Boccara, 1995; Bovy, 2009; Prato et al., 2012).

However, the way individual drivers derive their actual consideration set of route alternatives, and factors affecting the size and composition of these choice sets is a complex and ongoing research issue (Schuessler & Axhausen, 2009). In reality, individuals' choice sets are dependent on objective constraints, such as route attributes (e.g. maximum number of turns, number of traffic signals, etc.), as well as subjective criteria, such as individuals' attitudes, perceptions and experiences (e.g. spatial abilities, safety concerns, etc.) (Ben-Akiva & Boccara, 1995). For instance, some drivers may prefer scenic routes and the comfort of their car to enjoy their trip, or might prefer to avoid local routes and take highways to save time. Consequently, the formation of the consideration choice set and the process of choice selection may be very different from one individual to another and may depend on a wide range of factors, ranging from route attributes and accessibility to the road network to sociodemographic and behavioural factors.

This research aims to elaborate on this aspect of route choice modelling by observing drivers' actual consideration sets and analysing factors affecting them. The observation of drivers' consideration sets provides detailed information regarding their route selection behaviour and factors affecting such process (Bovy, 2009). It is worth mentioning that in this study the total number of declared alternatives is considered as the respondents' revealed consideration set.

Considered sets of route alternatives are mostly latent to the analyst and are rarely observed (Hoogendoorn-Lanser et al., 2005; Prato et al., 2012). This may be mostly due to the fact that collecting pertaining information is excessively time-consuming, increases the burden on respondents, and decreases response rates. To the authors' knowledge, a detailed study of drivers' revealed consideration sets has never been performed in the past. We present a web-based survey, designed to collect information on drivers' revealed consideration set of route alternatives. Then a comprehensive descriptive analysis is performed to shed light on factors affecting its size. The data contains information regarding drivers' sociodemographic and socioeconomic characteristics, route characteristics, as well as driving preferences and attitudes affecting their choices. The survey was conducted in the Greater Montreal Area and information on 506 drivers declaring 988 route alternatives was collected.

The remainder of the paper is structured as follows. First we describe our survey design and data collection methodology. Next we present a detailed descriptive analysis of factors affecting the size of drivers' consideration set. Then, we further discuss our findings and distinguish between the observed types of behaviour breeding different sizes of consideration sets. Finally, we underline the study's limitations and propose further research directions.

8.2 Survey

Data has been collected through a revealed preference web-based survey, designed to address the following three main purposes: i) to observe and characterize drivers' consideration set of route alternatives ii) to observe drivers' revealed route choices and iii) to identify behavioural and attitudinal factors affecting their route choice decisions and the formation of their considered choice sets. A detailed description of the survey, followed by a thorough review of the implementation steps is presented in this section. Then, a concise descriptive analysis of the data set clarifies the composition of the studied sample.

8.2.1 Survey Design

The employed survey consists of 6 separate sections. In the first and last sections, *Profile* and *End*, respectively, typical sociodemographic and socioeconomic data, such as age, gender, educational attainment, type of work, salary, household size, and number of cars in the household, is collected. Although the aim of this study is not to ascertain the representativeness of the sample, collecting these data provides the possibility of comparing the sampled population with the reference population (Ory & Mokhtarian, 2005). Moreover, this also provides a mean to segment the population based on important sociodemographic and socioeconomic factors affecting route choices, under the assumption that they differ based on population segments (Alizadeh, Farooq, Morency, & Saunier, 2017a).

In the second section, *Home*, participants provide their home addresses or pinpoint its location on a geographical map. This section explores factors such as familiarity with the road network around home locations, accessibility to the road network, transit services, and land use.

In the third section, *Trips*, respondents specify on a geographical map, the destination point to which they drive most frequently. Instances of these places can be work place, shopping malls, parents' place, school, etc. This section investigates several aspects of the declared trip, such as its purpose, the consulted information before making the trip and on the way, the familiarity with the road network around the destination location, and the travel frequency. Respondents are also asked to specify factors affecting their route choices as well as the number of route alternatives considered for the declared trip.

The fourth section, *Routes*, investigates route alternatives considered by respondents for their predefined trip. At first, they are asked to specify their considered routes on a geographical map by adjusting (dragging) an automatically generated route connecting the predefined origin and destination points. Then several questions are asked to better understand and characterize respondents' choices. These questions mostly focus on the effect of different factors on drivers' decisions, such as the amount of toll, time of day, weather conditions, drivers' perception of travel time and its reliability, safety, traffic conditions, scenery, and the number of traffic lights.

Finally, in order to collect information on behavioural and attitudinal variables affecting drivers' route choices, a list of different statements is provided to drivers in the fifth section, entitled *Preferences*. These statements are based on psychometric indicators chosen to reveal drivers' attitudes and perceptions towards choosing a route (Atasoy et al., 2013; Ory & Mokhtarian, 2005; Vredin Johansson et al., 2006). Statements are designed based on some behavioural assumptions on drivers' attitudes, in order to reveal the most important latent behavioural variables affecting drivers' route choices (Bierlaire, 2016). Respondents were asked to specify their level of agreement to the statements on a five-point Likert scale (Likert, 1932) ranging from total agreement to total disagreement. This section is designed to collect information on indicators, such as drivers' preferences towards riding on freeways or safer roads, willingness to try new routes, and openness to change route in case of accident or congestion. These indicators can be used to identify behavioural variables affecting drivers' consideration set of route alternatives (Atasoy et al., 2013; Raveau et al., 2010).

8.2.2 Survey Implementation

The survey only targets drivers residing and driving in the Greater Montreal Area, which covers an area of around 9840 square kilometers and contains a population of roughly 4 million inhabitants (Transport, 2013). It is a bilingual region with both French and English speaking populations; hence, the survey was prepared in both languages.

A web-based interface has been adopted to mitigate the implementation cost and obtain high resolution data (Bourbonnais & Morency, 2013). In order to minimize its complexity, geographical map interfaces were adopted for questions where respondents had to specify the origin and destination points of their trip, as well as their considered route alternatives. Graduate students of the Transportation Research Group of Polytechnique Montreal took part in a pilot test in February 2017, and the revised version of the survey was launched in March 2017.

To disseminate the survey, three means have been adopted, namely i) graduate students, postdocs, faculty members, and staff of Polytechnique Montreal, ii) social media, such as Facebook, LinkedIn, etc., and iii) a list of email addresses from volunteer participants who previously agreed to participate in surveys conducted by the Mobility Chair of Polytechnique Montreal.

8.2.3 Survey Participants

A total number of 532 respondents completed the survey over a period of two months, from which 26 were discarded due to geographical constraints and validation issues. The remaining 506 interviews reporting a total number of 988 routes were used for all the analysis reported in this paper. Table 8.1 provides comprehensive details regarding the respondents' sociodemographic and socioeconomic characteristics. It should be noted that the sample includes mainly young and middle aged full time workers with a university level of education. The skewness of the sample may be partly because the survey was disseminated among scholars, faculty members, and staff of Polytechnique Montreal. Moreover, the prevalence of young participants explains to some degree the higher frequency of smaller households.

Table 8.1: Characteristics of the Survey Participants

Variable categories	N	%	Variable categories	N	%
Gender			*Household size*		
Male	280	55.3	1	89	17.6
Female	226	44.7	2	207	40.9
Age (years old)			3	95	18.8
Young (15 to 39)	287	56.7	4	84	16.6
Middle age (40 to 59)	181	35.8	+5	31	6.1
Old (more than 60)	38	7.5			
Occupation			*Income (Thousand CAD per capita)*		
Full time worker	368	72.7	Low (Less than 35)	134	26.5
Partial time worker	37	7.3	Medium (35 to 75)	197	38.9
Student	61	12.1	High (more than 75)	79	15.6
Retired	24	4.7	Not declared	96	19
House-wife/husband	6	1.2			
Other	10	2	*Household car number*		
Education			0	108	21.3
None	0	0	1	275	54.3
Less than university	61	12.1	2	100	19.8
University	438	86.6	+3	23	4.6
Other	7	1.4			

8.3 Analysis and Results

To assess the effect of factors affecting the size of drivers' consideration set, individuals have been classified based on their number of declared alternatives. Pilot study suggested a maximum limit of five alternatives to be declared by any participant. Statistical analysis suggested that the classification over three clusters provides the best fit over the data. The initial clusters were created based on the number of considered alternatives, namely individuals reporting one alternative, two alternatives, and three or more alternatives. However, further analysis clarified that two different types of behaviours were observed among individuals having declared two alternatives (i.e. second cluster); namely, individuals excessively preferring one of the declared alternatives over the other one and behaving more similarly to the first cluster, and those behaving more similarly to the third cluster by fairly considering both declared alternatives. Therefore, it proved more logical and interpretable to further divide the second cluster, based on the usage frequencies of declared alternatives, in order to distinguish between their distinct behaviours. The resulting clusters are described below:

- Cluster 1 consists of 171 individuals considering only one route alternative.
- Cluster 2 includes 54 individuals declaring two alternatives and having one dominant alternative. A dominant alternative is defined as being chosen more than 80 % of the time.
- Cluster 3 contains 169 observations with two declared and frequently considered alternatives (i.e. declared routes are chosen more than 20 % and less than 80 % of the time).
- Cluster 4 incorporates a total of 112 individuals, considering three or more alternatives.

These pre-defined clusters are used in all the following analysis. We classified the studied factors into six broad categories, and separately assessed the relationship of each category with the size of the declared consideration sets. These categories include *personal attributes*, *declared factors*, *behavioural indicators*, *incentives*, *awareness determinants*, and *spatial, temporal and environmental components*.

To assess the effect of these attributes on the size of drivers' consideration choice set, the "Test Value" (TV) criterion has been used. The following formulas are used to calculate TV for continuous (t_c), and discrete (t_d) values (Lebart et al., 2000):

$$t_c = \frac{\mu_g - \mu}{\sqrt{\frac{n - n_g}{n - 1} \times \frac{\sigma^2}{n_g}}} \tag{8.1}$$

$$t_d = \frac{n_{jg} - \frac{n_g \times n_j}{n}}{\sqrt{\frac{n - n_g}{n - 1} \times \left(1 - \frac{n_j}{n}\right) \times \frac{n_g \times n_j}{n}}} \tag{8.2}$$

where μ and μ_g are attributes' means in the cluster and group, respectively; n and n_g denote the size of the cluster and the group, respectively; σ^2 represents the attribute variance in the cluster; and n_{jg} is the number of observations corresponding to the discrete attribute j in cluster g.

Moreover, the variation of TV values across the clusters has been visualized by sparklines in the following tables, where green and blue sparklines represent monotonically increasing and decreasing effects from cluster one to cluster four, respectively, and yellow sparklines represent non-monotonic changes across clusters.

8.3.1 Personal Attributes

The effect of personal attributes on the size of drivers' consideration set is illustrated in Table 8.2. It is noted that almost none of the personal attributes are monotonic across clusters, i.e. either increase or decrease from Cluster 1 to Cluster 4. This may partly be due to the fact that the sample is skewed towards young workers with a university degree, and older retired individuals, students, and less educated segments of the population are underrepresented. However, it can be inferred that middle-aged workers with high income, and students with average income tend to consider a higher number of alternative routes. Taking into account the higher number of cars per household in these clusters, it can be argued that these individuals may be more used to driving and have more driving experience. On the contrary, older retired individuals with an average salary and younger drivers with low income consider fewer alternatives. Also, the lower number of cars per household in these clusters may imply the lower usage of car, hence a lower driving experience.

Table 8.2: The Effect of Personal Attributes on the Size of Drivers' Consideration Set

Attributes	TV				Sparkline
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	
Gender					
Male	0.26	0.90	-0.85	0.01	
Female	-0.26	-0.90	0.85	-0.01	
Age					
Young	0.57	-0.18	0.03	-0.54	
Middle-aged	-1.73	-0.66	0.37	2.04	
Old	2.05	1.53	-0.71	-2.66	
Occupation					
Full time worker	0.13	-0.41	-0.61	0.85	
Partial time worker	0.90	-1.08	0.23	-0.49	
Student	-1.33	0.66	0.76	0.16	
House-wife/husband	-0.02	1.81	-0.87	-0.32	
Retired	1.72	-1.06	-0.01	-1.16	
Income (1000 CAD per capita)					
Low (less than 35)	-0.49	0.88	-0.59	0.57	
Average (35 to 75)	1.81	-0.69	1.39	-3.12	
High (more than 75)	-0.44	-0.96	-1.40	2.80	
Education					
Less than university	-0.18	0.66	-0.97	0.82	
University	-0.01	-1.16	1.02	-0.30	
Household					
Size	-0.83	1.89	0.56	-1.09	
Car number	-2.77	-0.14	0.23	3.01	

8.3.2 Declared Factors

Participants were asked to specify a maximum of five factors affecting their route choices. The relationship of these factors with the size of drivers' consideration set of route alternatives is investigated in this section. The number of observations and the level of significance of these factors with respect to each cluster is reported in Table 8.3.

We notice that the level of familiarity, safety, habit, and distance, decrease monotonically from Cluster 1 to Cluster 4, whereas the level of congestion, construction sites, traffic lights, and turns, increase monotonically. The behavioural interpretation of these monotonic changes is that individuals considering fewer alternatives prefer to consider shorter (in terms of distance), safer, and routes which they are more accustomed to. Moreover, they are less willing to consider alternative routes in order to avoid congestion, construction sites, traffic lights and higher number of turns. On the contrary, individuals willing to avoid congestion, construction sites, traffic lights, and multiple turns consider more route alternatives, and factors such as familiarity, habit, safety and less travel distance play a less important role in their decisions. Furthermore, although their effects are not monotonic across clusters, it is observed that scenic routes are more attractive to those considering fewer alternatives and travel time reliability is more important to those considering a higher number of alternatives. Also, individuals in Cluster 4 are more open to pay tolls and more determined to minimize their travel time by considering more alternative routes.

Table 8.3: The Relationship of Factors Affecting Drivers' Route Choices with the Size of Their Consideration Set of Route Alternatives

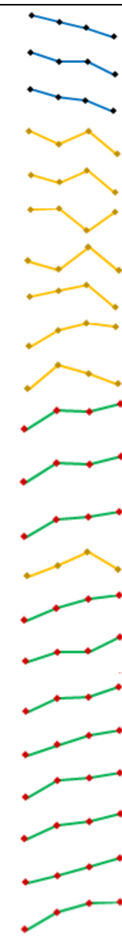
Attributes	N	TV				Sparkline
		Cluster 1	Cluster 2	Cluster 3	Cluster 4	
Higher familiarity	105	3.36	1.34	-2.10	-2.44	
Habit	175	1.94	1.00	-0.48	-2.41	
Higher safety	50	2.23	0.32	-1.17	-1.46	
Less distance	199	1.68	1.11	-1.24	-1.32	
More parking	56	0.92	-0.91	0.69	-1.16	
Higher speed limit	58	0.71	1.27	-1.59	0.05	
Less time	421	0.68	0.41	-2.92	2.23	
Better scenery	37	0.54	1.68	-1.21	-0.49	
Better pavement quality	37	0.18	1.13	-1.94	1.15	
Less toll	10	-0.26	-0.07	-0.23	0.60	
Higher travel time reliability	201	-0.37	-1.31	0.74	0.55	
Less turns	23	-1.25	-0.31	0.60	0.98	
Less traffic lights	109	-1.56	-0.22	0.59	1.27	
Fewer construction sites	87	-4.08	-0.49	1.48	3.33	
Less congestion	235	-4.97	-0.02	1.79	3.64	

8.3.3 Behavioural Indicators

Behavioural indicators and their effects on the size of drivers' consideration choice sets are illustrated in Table 8.4. It can be inferred that drivers considering more alternatives are more inclined to drive on highways and routes with higher speed limits. They prefer to have a more reliable travel time, and avoid traffic lights and construction sites. Also, they have a good sense of direction and take advantage of their reliable memories to remember routes they have taken a single time. Since they try to minimize their travel time, they would prefer taking longer routes rather than remaining stuck in traffic jams on shorter ones, and are consequently more prone to change route in case of accident or unexpected congestions. Moreover, these individuals are more likely to try new routes and are constantly in search of new shortcuts. Furthermore, it can be noticed that factors such as habit of taking the same route, scenery along the route, avoiding trucks and narrow lanes, and number of turns do not play a role in the formation of their consideration set.

Table 8.4: The Relationship of Behavioural Indicators with the Size of Drivers' Consideration Set of Route Alternatives

Attributes	Description	TV				Sparkline
		Cluster 1	Cluster 2	Cluster 3	Cluster 4	
HAB	Taking the same route repeatedly	5.77	1.87	-1.43	-6.34	
TRK	Avoiding routes with high number of trucks	2.49	0.09	0.24	-3.18	
DIR	Preferring more direct routes	1.51	0.30	-0.15	-1.78	
SCN	Preferring scenic routes	1.29	-0.76	1.24	-2.32	
NR_LN	Avoiding narrow lanes	1.18	-0.81	2.42	-3.49	
LOC	More willing to take local routes	0.27	0.34	-0.66	0.19	
TOL	Willing to pay	0.02	-0.52	0.86	-0.60	
RT_PLN	Taking suggested routes by route planners	-0.16	0.43	1.03	-1.31	
SPD	Preferring higher speed limits	-1.39	0.17	0.85	0.49	
LNDMRK	Using landmarks to recall new routes	-1.21	1.66	0.63	-0.58	
FRW	More willing to take freeways	-1.42	-0.03	0.67	0.87	
DRV_STCK	Preferring longer routes rather than being stuck in traffic jams on shorter routes	-1.42	0.42	0.27	1.00	
TT_REL	Having more reliable travel time	-1.87	0.26	0.64	1.21	
PAV_QLT	Importance of pavement quality	-1.53	0.07	1.93	-0.50	
SHRTCT	Searching for route shortcuts	-1.94	-0.27	0.88	1.41	
CONST	Avoiding construction sites	-2.29	-0.31	-0.12	2.98	
TT_MIN.	Minimizing travel time	-2.60	0.09	0.43	2.40	
SOD	Having a good sense of direction	-2.83	-0.79	1.37	2.26	
LGHT	Avoiding traffic lights	-3.41	0.45	1.20	2.19	
MEM	Having a good memory	-3.01	0.10	0.78	2.48	
NEWRT	Willingness to try new routes	-4.10	-1.67	1.49	4.22	
RT_CHNG	Changing route in case of accidents	-4.92	0.05	2.47	2.75	



Since almost all the above-mentioned factors are monotonic across clusters (except for scenery and narrow lanes), the behaviour of individuals considering fewer number of route alternatives is essentially the opposite of what has been described above. In other words, individuals considering fewer number of route alternatives mostly tend to follow their usual routes and are less willing to try new ones, find shortcuts, avoid construction sites, or even change route while facing unexpected traffic jams or accidents. This may partly be due to their weaker memorizing ability and sense of direction. These cautious drivers try to avoid the presence of trucks, routes with narrow lanes, multiple turns, and higher speed limits. As a final remark, it can be noted that less determined drivers of Clusters 2 and 3 are more frequent users of route planners.

8.3.4 Incentives

In this section, drivers' trip purpose and their incentives to choose car as their principal transportation mean for the declared trip have been investigated. Results are illustrated in Table 8.5.

Table 8.5: The Relationship of Trip Purpose and Incentives with the Size of Drivers' Consideration Set of Route Alternatives

Attributes	N	TV Cluster 1	TV Cluster 2	TV Cluster 3	TV Cluster 4	Sparkline
Trip purpose						
Work	204	-1.33	-1.40	0.55	1.93	
School	33	-0.82	-0.89	0.37	1.17	
Shopping	91	1.28	0.48	-0.10	-1.71	
Recreation	127	1.53	1.47	-0.52	-2.25	
Drive someone	34	-1.68	-0.36	0.62	1.48	
Services	17	0.65	0.95	-1.40	0.14	
Driving purpose						
Enjoy driving	78	-1.65	-0.92	-0.01	2.59	
Habit	16	1.93	0.24	-0.72	-1.55	
Free parking	42	-0.75	0.79	0.33	-0.11	
TT reliability	112	0.26	0.02	-0.55	0.31	
Comfort	106	1.65	0.95	-2.17	-0.12	
Time saving	301	-0.33	-0.62	-1.63	2.69	
Large objects	161	-1.29	0.56	1.26	-0.38	
Other people	130	-1.70	0.37	0.34	1.28	
Other activities	193	-0.24	-0.18	-0.09	0.50	

It can be noticed that drivers going to work, school, and those driving someone, consider higher number of alternative routes, and those driving for shopping and recreational activities are more willing to stick to their preferred route and consider fewer number of alternatives. This might be related to the fact that work and study trips mostly occur during peak hours and congested periods. With respect to driving purposes, on one hand, drivers who enjoy driving, drive to save time and spend lesser time on the road, or drive other people are more prone to consider various alternatives. On the other hand, drivers who drive out of habit and appreciate the comfort of their cars, consider fewer number of route alternatives.

8.3.5 Awareness Determinants

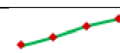
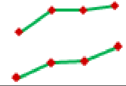
Attributes related to driving experience, familiarity with the road network, and consulted information are explored as determinants of drivers' state of awareness. Table 8.6(a) describes these determinants and Table 8.6(b) presents their relationships with the size of drivers' consideration set. It is noteworthy that eight out of the nine attributes show a monotonic behaviour across clusters. Accordingly, frequent drivers who are more experienced and consult traffic information, before and during the trip, consider a higher number of alternatives. In general, they have resided longer in their current home address and are consequently more familiar with the road network around it (their origin point). They also show a higher familiarity with the road network around their destination points, and in general with the road network of Montreal. Although the duration of residing in Montreal (RES_MTL) does not show a monotonic effect across clusters, it can be noticed that individuals in Cluster 4 have lived longer in Montreal, which probably rendered them more familiar with Montreal's road network and consequently more willing to try new routes.

Table 8.6: The Relationship of Awareness Determinants with the Size of Drivers' Consideration Set of Route Alternatives.

(a) List of Awareness Determinants

Attribute	Description
DRV_EXP	The percentage of respondents holding a driving license
TRIP_FREQ	Frequency of the declared trip per week
INF_BFR	Traffic information consulted before making the trip (internet, media = 1, none = 0)
INF_AFT	Traffic information consulted on the way (internet, media = 1, none = 0)
RES_HM	Duration that they have been living at the same address
RES_MTL	Duration that they have been residing in Montreal
ORG_FAM	Familiarity with the road network around the origin of the trip
DEST_FAM	Familiarity with the road network around the destination of the trip
MTL_FAM	Familiarity with the road network of Montreal

(b) Relationship of Awareness Determinants

Attributes	TV				Sparkline
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	
DRV_EXP	-1.00	-0.38	0.42	0.94	
TRIP_FREQ	-1.39	-1.09	-0.24	2.67	
INF_BFR	-2.67	-1.40	-0.22	4.33	
INF_AFT	-2.86	-1.11	0.57	3.44	
RES_HM	-1.20	-0.39	0.43	1.16	
RES_MTL	0.64	0.08	-3.32	2.98	
ORG_FAM	-2.12	0.63	0.70	1.15	
DEST_FAM	-3.69	1.09	1.14	2.10	
MTL_FAM	-2.01	0.07	0.19	2.02	

8.3.6 Spatial, Temporal and Environmental Components

In this section, we study the significance of spatial, temporal and environmental variables (presented in Table 8.7(a)) on the composition of drivers' consideration set (see Table 8.7(b)). It seems that drivers traveling away from CBD consider more route alternatives than drivers driving towards the CBD. It can also be inferred that individuals who have a flexible arrival time at work / school and travel on peak hours consider more route alternatives and are less sensitive to meteorological changes.

Table 8.7: The Relationship of Spatial, Temporal and Environmental Variables with the Size of Drivers' Consideration Set of Route Alternatives.

(a) List of Variables

Attribute	Description
LEN	The travelled distance
BRD_LEN	The bird fly distance between origin and destination points
ORG_CBD	The bird fly distance between the origin point and the Central Business District
DEST_CBD	The bird fly distance between the destination point and the Central Business District
TO_CBD	The direction of the trip (towards CBD = 1, away from CBD = 0)
FLX_ARV	Flexible arrival time at work or school (Yes = 1, No = 0)
PH	Peak hour (6 a.m. to 9a.m. and 3 p.m. to 7 p.m.) departure time (peak = 1, off-peak =0)
WTHR	The effect of weather on route choice (Yes = 1, No =0)

(b) Relationship of Variables

Attributes	TV				Sparkline
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	
LEN	0.42	0.39	-0.90	0.25	
BRD_LEN	0.76	0.53	-1.35	0.27	
ORG_CBD	1.97	-1.13	-0.44	-0.90	
DEST_CBD	2.12	1.33	-1.22	-2.02	
TO_CBD	-1.14	-1.58	1.12	1.20	
FLX_ARV	-2.15	-0.50	1.28	1.37	
PH	-1.40	0.33	0.29	1.01	
WTHR	2.78	-1.62	0.14	-2.11	

8.4 Summary and Conclusions

In a two-step choice modelling approach, the first stage consists of the formation of individuals' consideration set of alternatives, while the final choice from the considered set is performed in the second stage. The importance of delimiting individuals' consideration choice sets is well supported in the literature. It has been argued that the variability in the size of the choice set greatly affects models' estimation and prediction abilities (Bliemer & Bovy, 2008; Geda, 2014; Peters et al., 1995; Prato & Bekhor, 2006, 2007b; Schuessler & Axhausen, 2009; Swait & Ben-Akiva, 1987a). In route choice studies, consideration sets, their characteristics and factors affecting their sizes are usually unknown. Hence, various deterministic and probabilistic algorithms are used to generate a random number of alternatives per observation. Identifying the factors affecting the size and composition of drivers' actual consideration sets, is still an ongoing research problem. This paper reports on an empirical investigation of this issue using drivers' observed consideration sets.

Data was collected through a web-based survey, designed to collect detailed data on drivers' consideration set of route alternatives and factors affecting their compositions. The studied dataset is composed of 506 respondents residing and travelling in the Greater Montreal Area. A total of 988 route alternatives have been declared. Studied factors have been classified into six broad categories, namely *personal attributes*, *declared factors*, *behavioural indicators*, *incentives*, *awareness determinants*, and *spatial, temporal and environmental components*. Four clusters were defined based on the number of considered alternatives and the relationship of these factors with each cluster was studied. The behavioural interpretation of each cluster is concisely presented below:

Cluster 1 – Determined Cautious Drivers

Individuals in this cluster consider only one alternative for their declared trip. The presence of older retired people with an average income and fewer cars per household is highlighted in this cluster. People in this cluster seem to rely more on their habits and previous experiences, and prefer to take their usual route, to which they are more accustomed. They tend to take safer routes by avoiding trucks and routes with narrow lanes, and taking routes with lower speed limits. They more often use their cars for shopping and recreational purposes and like to enjoy their trips through the comfort of their cars and by taking more scenic routes. They are less interested in minimizing their travel times and prefer to take more direct routes with shorter travel distances. They are usually less experienced in driving, and less familiar with the road network around their origins, destinations, and Montreal in general. They barely consult traffic information, which may be partly because they are less affected by construction sites and congestion, and are less willing to change their routes to try new ones.

Cluster 2 – Biased Habitual Drivers

Individuals declaring two alternatives and favoring one of them more than 80 % of the time are classified in this cluster. It is characterized by a larger share of low income respondents, old individuals, and housekeepers. In general, their behaviour is similar to those in cluster one. However, compared to Cluster 1, the effect of various factors on their behaviour is mostly lower. These drivers are less determined to always use the same route and are more open to try new ones. As they become more familiar with the road network (around their origins, destinations, and Montreal in general), further experienced and more informed, the effect of habit becomes less prominent in their decisions. While they still prefer to use their usual route, they are less tolerant towards getting stuck in traffic jams and more prone to choose routes with reliable travel times.

Cluster 3 – Middling Impartial Drivers

This cluster includes drivers who consider two alternatives and are not biased towards any of them. Middle-aged individuals, students, and respondents with an average income are more represented in this cluster. Individuals in this cluster are more open-minded towards considering more route alternatives and more willing to explore new routes. Factors such as familiarity, habit and safety become less important to them and trying new routes and shortcuts gain importance in their decisions. Compared to previous clusters, they are more familiar with the road network and pay attention to traffic information more often.

Cluster 4 – Swayable Conscious Drivers

Individuals who consider three or more alternatives are classified in this cluster, which is highlighted by the presence of middle-aged workers with high income. These experienced drivers mostly live in highly motorised households. They usually drive to work and school and are very familiar with the road network around their origins, destinations, and Montreal in general. Since they drive to save time and spend less time on the road, they often consult traffic information (both before their trips and on their way) to identify construction sites and avoid congestion. They also seek to minimize their travel times by choosing more reliable routes and avoiding traffic lights. They have a keen sense of direction and a sharp memory, which make them more confident to explore new routes and try new shortcuts. Moreover, they are not distressed by the presence of trucks, higher speed limits and narrow lanes.

According to Horowitz and Louviere (1995), the observed consideration set of alternatives can improve choice model's estimation efficiency by providing information about travellers' preferences. Moreover, this information can also be used to improve models' prediction efficiency. Accordingly, the probability distribution of the random component of the utility function, conditional on the consideration choice set, can be calculated and used instead of the type I extreme value distribution, to calculate the choice probabilities for prediction purposes (Horowitz & Louviere, 1995). Although the studied factors and the premeditated clusters may not explain all the variations and stochasticity leading to diverse sizes of consideration sets, they shed light on the relationship of various attributes with the size and composition of the choice set.

A complementary study direction could be the study of choice set compositions, regarding alternative characteristics such as the overlapping portion between the declared alternatives, or the similarity between considered alternatives and generated alternatives using shortest path route generation techniques. Moreover, key factors and behavioural traits affecting drivers' choice selection behaviour can be embedded in the non-compensatory choice set formation step to generate and select more behaviourally realistic route alternatives.

Acknowledgments

The authors would like to express their gratitude to Pierre-Léo Bourbonnais for the design of the survey framework as well as the partners of the Mobilité Research Chair of Polytechnique Montréal for their financial support, providing access to data for research purposes and their help in the dissemination of the survey.

References

Alizadeh, H., Farooq, B., Morency, C., Saunier, N., 2017. Classifying Behavioral Dynamics of Taxi Drivers Route Choices Using Longitudinal GPS Data.

Atasoy, B., Glerum, A., Bierlaire, M., 2013. Attitudes towards mode choice in Switzerland. disP - The Planning Review 49, 101-117.

Ben-Akiva, M., Boccara, B., 1995. Discrete choice models with latent choice sets. International Journal of Research in Marketing 12, 9-24.

Bierlaire, M., 2016. Estimating choice models with latent variables with PythonBiogeme, EPFL.

Bliemer, M., Bovy, P., 2008. Impact of route choice set on route choice probabilities. Transportation Research Record: Journal of the Transportation Research Board, 10-19.

Bourbonnais, P.-L., Morency, C., 2013. Web-Based Travel Survey: A Demo, in: Johanna Zmud, M.L.-G., Marcela Munizaga, Juan Antonio Carrasco (Ed.), Transport Survey Methods: Best Practice for Decision Making, pp. 207-224.

Bovy, P.H., 2009. On modelling route choice sets in transportation networks: a synthesis. Transport reviews 29, 43-68.

Geda, A., 2014. Effect of Choice Set Composition on Route Choice Models. University of Florida.

Hoogendoorn-Lanser, S., van Ness, R., Bovy, P.H., 2005. Path Size and Overlap in Multi-modal Transport Networks, Transportation and Traffic Theory. Flow, Dynamics and Human Interaction. 16th International Symposium on Transportation and Traffic Theory.

Horowitz, J.L., Louviere, J.J., 1995. What is the role of consideration sets in choice modelling? International Journal of Research in Marketing 12, 39-54.

Lebart, L., PIRON, A., Lebart, M., Morineau, A., Piron, M., 2000. Statistique exploratoire multidimensionnelle. Dunod.

Likert, R., 1932. A technique for the measurement of attitudes. Archives of psychology.

Nadine Schuessler, I., Axhausen, K.W., 2009. Accounting for Route Overlap in Urban and Sub-urban Route Choice Decisions Derived from GPS Observations, 12th International conference on travel behavior research, Jaipur, India.

Ory, D.T., Mokhtarian, P.L., 2005. Don't work, work at home, or commute? Discrete choice models of the decision for San Francisco Bay Area residents.

Peters, T., Adamowicz, W.L., Boxall, P.C., 1995. Influence of choice set considerations in modeling the benefits from improved water quality. Water Resources Research 31, 1781-1787.

Prato, C., Bekhor, S., 2006. Applying branch-and-bound technique to route choice set generation. Transportation Research Record: Journal of the Transportation Research Board, 19-28.

Prato, C., Bekhor, S., 2007. Modeling route choice behavior: how relevant is the composition of choice set? Transportation Research Record: Journal of the Transportation Research Board, 64-73.

Prato, C.G., Bekhor, S., Pronello, C., 2012. Latent variables and route choice behavior. Transportation 39, 299-319.

Raveau, S., Álvarez-Daziano, R., Yáñez, M., Bolduc, D., de Dios Ortúzar, J., 2010. Sequential and simultaneous estimation of hybrid discrete choice models: some new findings. Transportation Research Record: Journal of the Transportation Research Board, 131-139.

Swait, J., Ben-Akiva, M., 1987. Empirical test of a constrained choice discrete model: mode choice in Sao Paulo, Brazil. Transportation Research Part B: Methodological 21, 103-115.

Transport, A.M.d., 2013. Enquête Origine-Destination 2013: La mobilité des personnes dans la région de Montréal-FAITS SAILLANTS, in: Montreal, A.M.d.T. (Ed.), Montreal.

Vredin Johansson, M., Heldt, T., Johansson, P., 2006. The effects of attitudes and personality traits on mode choice. Transportation Research Part A: Policy and Practice 40, 507-525.

CHAPTER 9 GENERAL DISCUSSION

The main contributions of this thesis are related to the behavioural enrichment of the two-stage RUM framework with sampling of alternatives. This first requires the definition of a considered choice set from which individuals make their final choices in the second stage. The application of the two-stage random utility maximization framework in route choice modelling gives rise to a particular challenge, namely the definition of realistic and representative choice sets. Another challenge of route choices modelling is to consider the complex correlation structure of route alternatives. The main objective of this thesis is to enhance the behavioural understanding of route choice decisions using drivers' underlying behavioural process of decision-making. Behaviourally elaborated models require customized programs and fast computers for estimation and necessitate well-tailored data collection methods. However, according to Walker (2001) these models provide better prediction abilities, correct for cognitive biases, verify behavioural hypotheses regarding the decision-making process, allow for a clearer behavioural interpretation than standard choice models, and hence, provide a benchmark to evaluate the performance of more parsimonious models.

In this regard, the main contributions of the thesis revolve around the three general components of the two-stage route choice modelling framework, also discussed in Chapter 3, namely the modelling framework, the data collection method, and the consideration set of route alternatives. Behavioral contributions to the modelling framework have been presented in detail in Chapters 4, 5, and 7. Three main ideas were discussed and studied:

- In Chapter 4, we followed the idea that drivers follow the hierarchical representation of space and that some prominent features of the route, i.e. anchor points, might affect their decisions. Although several studies have argued that anchor points influence route choice decisions (Couclelis et al., 1987; Golledge et al., 1985; Habib et al., 2013; Kaplan & Prato, 2012; Lynch, 1960; Prato & Bekhor, 2007a; Prato et al., 2012), route choice studies have mostly ignored their importance by only considering route level attributes, such as the total travel time and number of traffic signals. We studied the influence of bridges as anchor points on drivers' route choice decisions between the two islands of Montreal and Laval. We adopted a nested modelling approach to represent the space hierarchy and to incorporate the effects of anchor points and route level attributes at the same time. Moreover, considering anchor points also provides the possibility to capture the unobserved similarities of routes crossing the same anchor points, such as safety, scenery, driving comfort, etc. Results illustrated that the proposed nested modelling approach provides better model fits and outperforms the prediction abilities of comparative route-based choice models.

- Chapter 5 elaborates on the idea that individuals have different inclinations towards choosing a route between an origin and destination, which may form different types of decision-making behaviours. It has been discussed that behavioural differences mostly emerge from having different attitudes, preferences, and experiences. Since previous studies have shown that population stratification is effective in capturing the preference heterogeneity of the population, a representative classification of drivers' route choice decisions based on their actual choices has been investigated. In the presented study, it has been illustrated that observing taxi drivers' route choice decisions for a long period of time can shed light on their route choice behaviours, and that different types of operational strategies can be observed. Although GPS traces lack important explanatory variables, such as demographic attributes, attitudes and preferences, this study shed some light on the variation of taxi drivers' route choice behaviours and the possibility of classifying them.
- The third contribution, discussed in Chapter 7, revolves around the idea that choices are greatly influenced by: 1) latent traits and variables that cannot be directly observed and measured, such as attitudes, perceptions, and lifestyle preferences, and 2) the latent heterogeneity existing between different segments of the population. The study integrates an ICLV model into a LC model to explicitly capture the behavioural aspect of the choice process. The study implements the modelling framework to improve the understanding of frequent versus occasional drivers' route choice behaviour. The better fit and behavioural interpretation of the proposed modelling framework underscored the role of underlying behavioural constructs on drivers' route choice decisions.

Despite the appeal of behavioural modelling frameworks, their application in route choice studies remains rare, which can be mostly related to the fact that collecting behavioural data is cumbersome and time consuming. In this thesis, we proposed a data collection framework designed for behavioural route choice studies.

- The main aim of the proposed framework is to collect data reflecting the heterogeneity of individuals' preferences and the complex nature of drivers' decision-making process, without significantly increasing the respondent burden. Drivers residing and driving in the GMA have been targeted. After a three-month data collection period, the final dataset included the complete validated responses from 513 participants.

  We also looked at different types of questions, their completion times, and the rate of dropouts per question type and section. Statistics on various recruitment methods and survey completion percentage per hour of the day are also discussed. These paradata, which were available thanks to the adopted web-based interface, provide various potential benefits to tackle some of the main challenges facing survey developers, namely increasing the response rate, decreasing the risk of non-response bias, increasing response precision and minimizing the survey error, increasing the reliability and efficiency of the collected data, and reducing the overall cost of the survey. Paradata can be used to identify data collection problems, propose new data collection strategies, and determine a trade-off between data quality, cost and time (Nicolaas, 2011).

The last contributing area of this thesis concerns the first stage of route choice modelling in a two-stage modelling framework, namely the understanding of route choice sets.

- Since actual consideration sets of route alternatives are usually latent to the analyst and information concerning objective and subjective attributes affecting their size and composition is limited, we adopted the proposed survey framework to collect information on respondents considered choice sets. Although the studied factors in this study may not explain all the variations and stochasticity leading to diverse sizes of consideration sets, they shed light on the relationship of various attributes with the size of the choice set. The identification of key factors and behavioural traits affecting drivers' choice selection behaviour can be embedded in the non-compensatory choice set formation step to generate and select more behaviourally realistic route alternatives.

CHAPTER 10 CONCLUSION AND RECOMMENDATIONS

This section includes a summary of the studies presented throughout this thesis, research contributions, research limitations, and finally the perspectives for future works.

10.1 Summary

The work presented in the thesis revolves around several ways of improving the understanding of drivers' route choice decisions. These research efforts are presented in detail in five chapters. Here is a summary of the main points covered in these chapters.

10.1.1 Hierarchy of Space and the Role of Anchor Points

In Chapter 4, the effect of anchor points and space hierarchy in drivers' decision-making process has been investigated. Accordingly, the effect of bridges connecting Montreal to its Northern suburb, Laval, on taxi drivers' route choice decisions between these two riverside cities has been studied. These bridges face recurrent congestion and significant travel time variations, and even though they cover a small portion of the whole trip, they have a major impact on the experienced travel time. Therefore, they are considered as important points along the route, which might influence drivers' route choice decisions.

In order to incorporate the effect of space hierarchy in the drivers' decision-making process and to improve the behavioural aspect of route choice modelling, we have explored the application of a nested structure. First, a NL formulation has been proposed, in which upper nests represent bridges and lower nests consist of route alternatives crossing the respective bridges. Second, a nested LK with a factor analytic structure is specified. The EPS factor has been added to the deterministic part of the utility function to account for physical overlaps among routes crossing the same bridges. Routes crossing the same anchor point share unobserved components such as safety, scenery, driving comfort, etc., which is mainly because they share the same network and geographical characteristics. These unobserved similarities are captured through the nested structure and the factor analytic structure in NL and LK models, respectively.

To evaluate the estimation and prediction performance of these models, they have been compared to other route-based models, and findings revealed that the nested structures provided better model fits and prediction accuracies. This underscored the importance of considering the effect of anchor points in conjunction with route level attributes in route choice decisions.

10.1.2 Behavioural Classification

In the research effort, presented in Chapter 5, we studied taxi drivers' route choices to investigate possible types of route choice behaviours. We presumed that since taxi drivers have extensive driving experience, they develop different driving habits and behaviours, which breed different types of operating strategies. In other words, we hypothesised that factors affecting route choice decisions, such as preferences, experiences, information levels, and attitudes, are somehow correlated and can be classified to represent various types of route choice behaviours, hence, operating strategies.

For this purpose, we studied a longitudinal GPS dataset, tracking 1,746 taxi drivers making more than 22,000 trips over a period of one year. Accordingly, four categories of operating strategies have been found based on variations in trips made during days and nights, and between short trips and long trips.

Although it is not possible to encompass all variations of operating strategies based on route choice behaviours and GPS traces alone, due to the lack of some other explanatory variables, such as demographics and preferences, the main goal of this study was to shed light on the possibility of classifying drivers' decision-making behaviours based on their actual route choices. Apart from the fact that the understanding of these operating strategies helps to better comprehend urban traffic dynamics, which is very important to the city and transportation planners, the behavioural classification provides the possibility of estimating more behaviourally accurate route choice models.

10.1.3 Specialized Data Collection

In Chapter 6, we present the development and deployment of a general data collection framework adapted for behavioural route choice studies. The survey has been developed in six separate sections collecting information on drivers' sociodemographic and socioeconomic characteristics, their revealed route choices and their considered sets of route alternatives, as well as their perceptions and behavioural traits. Several validation criteria were defined for each question, and responses were required to comply with all the criteria in order to be approved and stored in the database. Respondents could advance to the next section only if they had answered all the questions in the previous section. By the end of the three-month data collection period, 843 individuals started the survey from which 539 (64 %) completed it, while the remaining 304 (36 %) dropped out at various points of the survey. The average completion time of the survey was around 16 minutes. The overall survey completion percentage of 64 %, and the small number of discarded interviews (4.8 %) suggests a successful implementation of the survey framework and the high quality of the collected data.

10.1.4 Behavioural Traits and Latent Heterogeneity

In Chapter 7, we proposed a route choice modelling framework, which incorporates the effect of latent behavioural constructs in conjunction with population segment heterogeneity. We applied the proposed modelling framework to compare the route choice behaviour of frequent versus occasional drivers. To properly incorporate the effect of segment heterogeneity and to distinguish between choice behaviours of the different classes of our sample population we used a LC model, in which we incorporated the role of the underlying attitudinal and behavioural traits using an ICLV model. Data has been collected through the revealed preference web-based survey described in the previous sub-section (see Chapter 6 for a thorough description). The modelling dataset included 225 drivers residing and driving in the GMA.

As expected, major behavioural traits have been observed among drivers, which were associated to different segments of the studied population, i.e. frequent and occasional drivers, affecting their route choice decisions. A latent class model without the ICLV component has also been estimated as a benchmark to compare the results. The same specification of the utility function and class-membership equations have been used in both models. Results demonstrated that the inclusion of behavioural traits in the LC model significantly improves its fit over the data.

10.1.5 Consideration Set of Route Alternatives

In order to better understand drivers' choice set formation process, we investigated 988 route alternatives, declared by 506 drivers, residing and driving in the Greater Montreal Area. Data has been collected through the web-based data collection framework presented in Chapter 6.

Then, the effect of six broad categories of factors on the size of drivers' consideration sets has been studied, including personal attributes, declared factors, behavioural indicators, incentives, awareness determinants, and spatial, temporal and environmental components. Accordingly, four different clusters were defined based on the number of considered alternatives and the relationship of these factors with each cluster was investigated.

Although the studied factors may not explain all the variations and stochasticity leading to diverse sizes of consideration sets, they shed light on the relationship of various attributes with the size and composition of the choice set. A better understanding of drivers' consideration sets' size and composition may significantly improve route choice models' estimation and prediction efficiency by providing information about travellers' preferences.

10.2 Research Contributions

The following contributions have been made throughout this thesis:

10.2.1 Hierarchy of Space and the Role of Anchor Points

The presented anchor-based nested modelling framework improves the behavioural aspect of route choice modelling by capturing the effect of space hierarchy and anchor points in conjunction with route level attributes. Moreover, the adopted NL and LK models are easily manageable and practical, even by considering many route alternatives crossing each anchor point. Moreover, the inclusion of multiple landmarks and anchor points, and the consideration of several forms of decision makers' preference heterogeneity and taste variation is easily manageable using the LK model, due to its flexible structure of the error term.

10.2.2 Behavioural Classification

This study shed some light on taxi drivers' different types of driving patterns and route choice strategies that are observable through a longitudinal route choice datasets. The incorporation of these categories in route choice models generally improves their estimation and prediction accuracy, by better capturing the existing heterogeneity among different segments of the population.

10.2.3 Specialized Data Collection

The data collection method proposed for route choice studies provides a general data collection framework to collect data on drivers' actual route choices, reveal drivers' consideration set of route alternatives, and identify important factors, including observable attributes and latent behavioural traits, affecting their decisions. The study also shed some light on participants' response behaviour and dropout rate, with respect to various types of questions used in the survey.

10.2.4 Behavioural Traits and Latent Heterogeneity

The proposed LC-ICLV modelling framework incorporates the effect of latent behavioural constructs in conjunction with population segment heterogeneity. The study also compares frequent versus occasional drivers and provides useful insights on that matter.

10.2.5 Consideration Set of Route Alternatives

Observing drivers' consideration set of route alternatives provides important insights on factors affecting their sizes and compositions. These influencing factors can be used in choice set generation techniques to provide more realistic choice sets, and hence improving choice model's estimation efficiency by providing information about travellers' preferences.

10.3 Research Limitations

Although the results obtained in this research effort have contributed to the existing literature on route choice modelling, some limitations are present. In chapters four to eight we have mentioned some limitations specific to each study. Some more general limitations are discussed below:

10.3.1 Data Availability

In some cases, the lack of necessary data represents major limitations that deserve special attention for future developments. For instance, in anchor-based models, including physical characteristics of anchor points (such as the travel time associated with them) can also be interesting and can provide useful insights on their effects on the attractiveness of an alternative. The inclusion of these factors is expected to enhance the models' estimation and prediction abilities. Moreover, in the presented behavioural classification of route choice patterns, a major limitation is the lack of personal information such as demographic and socio-economic characteristics of the participants, and their behavioural and attitudinal variables.

10.3.2 Population Representativeness

In research studies, it is recommended that the sampled population represents the target population. This is mostly to ensure that the sampled population consists of all relevant types of people and behavioural preferences, so that the findings from the study can be applied to the target population. The data collected through the web-based survey discussed in Chapter 6, has been used in studies presented in Chapters 7 and 8. A notable limitation is that it was not practically feasible to ensure a representative sample of the population due to lack of time, budget, and research necessities. This might have affected the results to be biased towards a particular segment of the population.

Also, the behavioural classification effort, presented in Chapter 5, is based on taxi drivers' route choice decisions and is not necessarily representative of the entire population of drivers.

10.4 Directions for Future Research

Considering the above-mentioned contributions towards enhancing the behavioural aspect of route choice modelling, this work has also opened doors to many research areas, some of which are presented below.

10.4.1 Studied Factors

To improve the findings of the presented work, one research direction could be the inclusion of other types of factors that could influence drivers' route choice behaviours. For instance, the inclusion of different attitudinal indicators, and more complex specifications of the class-membership functions may improve the estimation and prediction accuracy of the LC-ICLV model presented in Chapter 7. Moreover, the specification of a more complex utility function, including anchor points' characteristics in Chapter 4 may yield better model fits and prediction accuracies.

10.4.2 Time and space transferability

For future works, it would also be interesting to investigate the spatial and temporal transferability of the proposed modelling structures and behavioural classifications, presented in this research effort, for different datasets on similar case studies in different regions, countries and segments of the population.

10.4.3 Other Modelling Frameworks

In this work, we adopted a two-stage RUM framework to model drivers' route choice decisions, which requires the definition of a considered choice set or an alternative sampling procedure. In this work we did not analysed the sensitivity of the estimated parameters to the sampling size, which can be an extension of this thesis. Many other approaches are available, such as Recursive Logit (RL), proposed by (Fosgerau et al., 2013), which does not require the generation of a consideration set of route alternative. An interesting venue of research would be to incorporate the effects of anchor points and taste heterogeneity in terms of latent variables and latent classes in the RL modelling framework and to compare the results with the estimated modelling frameworks in this thesis.

10.4.4 Data Collection

In this work, we have presented a particular type of data collection framework, containing a specific number of questions and question types, using a web-based interface. In order to be able to better evaluate the efficiency and data quality of the proposed framework it should be compared with other route choice data collection frameworks, using different interfaces, with different lengths and types of questions. Another possible extension of this effort can be the integration of the proposed survey framework with other data collection mediums, such as smartphones and GPS devices, to compare declared and actual route choices.

10.4.5 Choice Set Composition

In this work, we studied the relation of various factors with the size of drivers' consideration sets of route alternatives. An extension to this work is the adoption of a more systematic way of capturing the effects of the studied factors, such as performing a clustering analysis or applying a Latent Class Analysis. Another complementary study direction is the study of choice set compositions, regarding alternative characteristics such as the overlapping portion between the declared alternatives, or the similarity between considered alternatives and generated alternatives using shortest path route generation techniques.

BIBLIOGRAPHY

Abdel-Aty, M., & Jovanis, P. (1997). *Using GIS Capabilities to Improve The Understanding Of Route Choice Behavior*. Paper presented at the Transportation Research Board, Washington D.C.

Abdel-Aty, M. A. (1994). Models of commuters' information use and route choice : initial results based on Southern California commuter route choice survey. *Transportation research record.*(1453).

Aitken, R., Conway, A., & Walton, M. (2012). Implementing Graphical Route Information Panel in the United States. *ITE Journal*.

Alizadeh, H., Bourbonnais, P. L., Morency, C., Farooq, B., & Saunier, N. (2017). *An Online Survey to Enhance the Understanding of Car Drivers Route Choices*. Paper presented at the International Conference on Transport Survey Methods, Montreal, Canada.

Alizadeh, H., Farooq, B., Morency, C., & Saunier, N. (2017a). *Classifying Behavioral Dynamics of Taxi Drivers Route Choices Using Longitudinal GPS Data*. Retrieved from

Alizadeh, H., Farooq, B., Morency, C., & Saunier, N. (2017b). On the role of bridges as anchor points in route choice modeling. *Transportation*. doi:10.1007/s11116-017-9761-7

Arifin, Z. N. (2012). *Route choice modeling based on GPS tracking data.* Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 20400, 2012.

Atasoy, B., Glerum, A., & Bierlaire, M. (2013). Attitudes towards mode choice in Switzerland. *disP - The Planning Review, 49*(2), 101-117. doi:10.1080/02513625.2013.827518

Azevedo, J., Costa, M. E. O. S., Madeira, J. J. E. S., & Martins, E. Q. V. (1993). An algorithm for the ranking of shortest paths. *European Journal of Operational Research, 69*(1), 97-106.

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of statistical psychology, 3*(2), 77-85.

Bartlett, M. S. (1951). A further note on tests of significance in factor analysis. *British Journal of statistical psychology, 4*(1), 1-2.

Bekhor, S., Ben-Akiva, M., & Ramming, S. (2002). Adaptation of logit kernel to route choice situation. *Transportation Research Record: Journal of the Transportation Research Board*(1805), 78-85.

Bekhor, S., & Prashker, J. (2001). Stochastic User Equilibrium Formulation for Generalized Nested Logit Model. *Transportation Research Record: Journal of the Transportation Research Board, 1752*(-1), 84-90. doi:10.3141/1752-12

Ben-Akiva, M. (1973). *Structure of passenger travel demand models.* (Ph.D.), Massachusetts Institute of Technology, Cambridge, MA.

Ben-Akiva, M., Bergman, M., Daly, A. J., & Ramaswamy, R. (1984). *Modeling inter-urban route choice behaviour.* Paper presented at the Proceedings of the 9th International Symposium on Transportation and Traffic Theory, VNU Press, Utrecht.

Ben-Akiva, M., & Bierlaire, M. (1999a). Discrete choice methods and their applications to short term travel decisions *Handbook of transportation science* (pp. 5-33): Springer.

Ben-Akiva, M., & Bierlaire, M. (1999b). Discrete Choice Methods and their Applications to Short Term Travel Decisions. In R. Hall (Ed.), *Handbook of Transportation Science* (Vol. 23, pp. 5-33): Springer US.

Ben-Akiva, M., & Bierlaire, M. (2003). Discrete choice models with applications to departure time and route choice. *Handbook of transportation science*, *32*.

Ben-Akiva, M., & Boccara, B. (1995). Discrete choice models with latent choice sets. *International Journal of Research in Marketing, 12*(1), 9-24.

Ben-Akiva, M., Bolduc, D., & Bradley, M. (1993). *Estimation of travel choice models with randomly distributed values of time*.

Ben-Akiva, M., Bolduc, D., & Walker, J. (2001). Specification, identification and estimation of the logit kernel (or continuous mixed logit) model. *Department of Civil Engineering Manuscript, MIT*.

Ben-Akiva, M., De Palma, A., & Isam, K. (1991). Dynamic network models and driver information systems. *Transportation Research Part A: General, 25*(5), 251-266. doi:http://dx.doi.org/10.1016/0191-2607(91)90142-D

Ben-Akiva, M., Walker, J., Bernardino, A. T., Gopinath, D. A., Morikawa, T., & Polydoropoulou, A. (2002). Integration of choice and latent variable models. *Perpetual motion: Travel behaviour research opportunities and application challenges*, 431-470.

Ben-Elia, E., Erev, I., & Shiftan, Y. (2008). The combined effect of information and experience on drivers' route-choice behavior. *Transportation, 35*(2), 165-177. doi:10.1007/s11116-007-9143-7

Ben-Elia, E., & Shiftan, Y. (2010). Which road do I take? A learning-based model of route-choice behavior with real-time information. *Transportation Research Part A: Policy and Practice, 44*(4), 249-264. doi:http://dx.doi.org/10.1016/j.tra.2010.01.007

Bierlaire, M. (2003). *BIOGEME: a free package for the estimation of discrete choice models.* Paper presented at the Swiss Transport Research Conference.

Bierlaire, M. (2015). BisonBiogeme 2.4: estimating a first model. *Series on Biogeme TRANSP-OR, 150720*.

Bierlaire, M. (2016). *Estimating choice models with latent variables with PythonBiogeme*. Retrieved from EPFL: http://biogeme.epfl.ch/documents.html#latent

Bierlaire, M., Bolduc, D., & McFadden, D. (2008). The estimation of generalized extreme value models from choice-based samples. *Transportation Research Part B: Methodological, 42*(4), 381-394. doi:http://dx.doi.org/10.1016/j.trb.2007.09.003

Bierlaire, M., & Fetiarison, M. (2009). *Estimation of discrete choice models: extending BIOGEME.* Paper presented at the Swiss Transport Research Conference (STRC).

Bierlaire, M., & Frejinger, E. (2005). *Route choice models with subpath components.* Paper presented at the Swiss Transportation Research Conference.

Bierlaire, M., & Frejinger, E. (2008). Route choice modeling with network-free data. *Transportation research part C: emerging technologies, 16*(2), 187-198.

Bliemer, M., & Bovy, P. (2008). Impact of route choice set on route choice probabilities. *Transportation Research Record: Journal of the Transportation Research Board*(2076), 10-19.

Bolduc, D., & Ben-Akiva, M. (1991). *A multinomial probit formulation for large choice sets.* Paper presented at the Proceedings of the 6th International Conference on Travel Behaviour.

Bourbonnais, P.-L., & Morency, C. (2013). Web-Based Travel Survey: A Demo. In M. L.-G. Johanna Zmud, Marcela Munizaga, Juan Antonio Carrasco (Ed.), *Transport Survey Methods: Best Practice for Decision Making* (pp. 207-224).

Bovy, P. (2009). On modelling route choice sets in transportation networks: a synthesis. *Transport reviews, 29*(1), 43-68.

Bovy, P., Bekhor, S., & Prato, C. G. (2008). *The Factor of Revisited Path Size Alternative Derivation.* Paper presented at the Transportation Research Board.

Bovy, P., & Stern, E. (1990a). *Route choice: Wayfinding in transport networks*: KLUWER ACADEMIC PUBLISHERS.

Bovy, P. H., & Stern, E. (1990b). Route choice: Wayfinding in transportation networks: Kluwer Academic Publishers, Dordrecht.

Bradley, M. (2013). Workshop Synthesis: Measuring the Influence of Attitudes and Perceptions *Transport Survey Methods: Best Practice for Decision Making* (pp. 643-646): Emerald Group Publishing Limited.

Bricka, S., Zmud, J., Wolf, J., & Freedman, J. (2009). Household Travel Surveys with GPS. *Transportation Research Record: Journal of the Transportation Research Board, 2105*(1), 51-56.

Bricka, S. G., Sen, S., Paleti, R., & Bhat, C. R. (2012). An analysis of the factors influencing differences in survey-reported and GPS-recorded trips. *Transportation research part C: emerging technologies, 21*(1), 67-88.

Brownstone, D., Bunch, D. S., & Train, K. (2000). Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B: Methodological, 34*(5), 315-338.

Cantillo, V., Arellana, J., & Rolong, M. (2015). Modelling pedestrian crossing behaviour in urban roads: A latent variable approach. *Transportation Research Part F: traffic psychology and behaviour, 32*, 56-67. doi:https://doi.org/10.1016/j.trf.2015.04.008

Cascetta, E., Nuzzolo, A., Russo, F., & Vitetta, A. (1996a). *A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks.* Paper presented at the Thirteenth International Symposium on Transportation and Traffic Theory, Pergamon, Lyon, France.

Cascetta, E., Nuzzolo, A., Russo, F., & Vitetta, A. (1996b). *A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks.* Paper presented at the Proceedings of the 13th International Symposium on Transportation and Traffic Theory.

Cascetta, E., & Papola, A. (2001). Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. *Transportation research part C: emerging technologies, 9*(4), 249-263. doi:http://dx.doi.org/10.1016/S0968-090X(00)00036-X

Cascetta, E., Russo, F., Viola, F. A., & Vitetta, A. (2002). A model of route perception in urban road networks. *Transportation Research Part B: Methodological, 36*(7), 577-592. doi:http://dx.doi.org/10.1016/S0191-2615(00)00041-2

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research, 1*(2), 245-276.

Chatterjee, K., Hounsell, N. B., Firmin, P. E., & Bonsall, P. W. (2002). Driver response to variable message sign information in London. *Transportation research part C: emerging technologies, 10*(2), 149-169. doi:http://dx.doi.org/10.1016/S0968-090X(01)00008-0

Chen, J. (2013a). *Modeling Route Choice Behavior Using Smartphone Data.* (Doctorat), Polytechnique Lausanne, Lausanne, Suisse.

Chen, J. (2013b). *Modeling route choice behavior using smartphone data.* (PhD), Polytechnique de Lausanne, Suisse.

Communauté métropolitaine de montréal, C. (2012). *An Attractive, Competitive and Sustainable Greater Montreal*. Montréal, Québec: Library and Archives Canada

Connors, R. D., & Sumalee, A. (2009). A network equilibrium model with travellers' perception of stochastic travel times. *Transportation Research Part B: Methodological, 43*(6), 614-624. doi:http://dx.doi.org/10.1016/j.trb.2008.12.002

Cools, M., Moons, E., & Wets, G. (2009). Assessing the Impact of Public Holidays on Travel Time Expenditure Differentiation by Trip Motive.

Couclelis, H., Golledge, R. G., Gale, N., & Tobler, W. (1987). Exploring the anchor-point hypothesis of spatial cognition. *Journal of Environmental Psychology, 7*(2), 99-122.

Couper, M. P., & Kreuter, F. (2013). Using paradata to explore item level response times in surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 176*(1), 271-286.

Cramer, D., & Howitt, D. L. (2004). *The Sage dictionary of statistics: a practical resource for students in the social sciences*: Sage.

Dalumpines, R., & Scott, D. M. (2017a). Determinants of route choice behavior: A comparison of shop versus work trips using the Potential Path Area-Gateway (PPAG) algorithm and Path-Size Logit. *Journal of Transport Geography, 59*, 59-68.

Dalumpines, R., & Scott, D. M. (2017b). Determinants of route choice behavior: A comparison of shop versus work trips using the Potential Path Area - Gateway (PPAG) algorithm and Path-Size Logit. *Journal of Transport Geography, 59*, 59-68. doi:http://dx.doi.org/10.1016/j.jtrangeo.2017.01.003

Damant-Sirois, G., Grimsrud, M., & El-Geneidy, A. M. (2014). What's your type: a multidimensional cyclist typology. *Transportation, 41*(6), 1153-1169. doi:10.1007/s11116-014-9523-8

de la Barra, T., Perez, B., & Anez, J. (1993). *Multidimensional path search and assignment.* Paper presented at the PTRC Summer Annual Meeting, 21st, 1993, University of Manchester, United Kingdom.

Dhakar, N., & Srinivasan, S. (2014). Route choice modeling using GPS-based travel surveys. *Transportation Research Record: Journal of the Transportation Research Board*(2413), 65-73.

Dhaker, N. S. (2012). *Route choice modeling using GPS data*: University of Florida.

Dia, H. (2002). An agent-based approach to modelling driver route choice behaviour under the influence of real-time information. *Transportation research part C: emerging technologies, 10*(5–6), 331-349. doi:http://dx.doi.org/10.1016/S0968-090X(02)00025-6

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische mathematik, 1*(1), 269-271.

Dill, J., & McNeil, N. (2013). *Four types of cyclists? Examining a topology to better understand bicycling behavior and potential.* Paper presented at the 92nd Annual Meeting of the Transportation Research Board, Washington, DC.

Dougherty, M. (1995). A review of neural networks applied to transport. *Transportation research part C: emerging technologies, 3*(4), 247-260. doi:http://dx.doi.org/10.1016/0968-090X(95)00009-8

Duan, Z., & Wei, Y. (2014). Revealing Taxi Driver Route Choice Characteristics Based on GPS Data. *CICTP 2014@ sSafe, Smart, and Sustainable Multimodal Transportation Systems*, 565-573.

Eagle, N., & Pentland, A. S. (2009). Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology, 63*(7), 1057-1066. doi:10.1007/s00265-009-0739-0

Elgar, I., Farooq, B., & Miller, E. (2009). Modeling location decisions of office firms: Introducing anchor points and constructing choice sets in the model system. *Transportation Research Record: Journal of the Transportation Research Board*(2133), 56-63.

Elgar, I., Farooq, B., & Miller, E. J. (2015). Simulations of firm location decisions: Replicating office location choices in the Greater Toronto Area. *Journal of Choice Modelling, 17*, 39-51. doi:http://dx.doi.org/10.1016/j.jocm.2015.12.003

Everitt, B. S., Stahl, D., Leese, M., & Landau, S. (2011). Cluster analysis.

Flötteröd, G., & Bierlaire, M. (2013). Metropolis–Hastings sampling of paths. *Transportation Research Part B: Methodological, 48*, 53-66.

Foo, P., Warren, W. H., Duchon, A., & Tarr, M. J. (2005). Do humans integrate routes into a cognitive map? Map-versus landmark-based navigation of novel shortcuts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 195.

Fosgerau, M., Frejinger, E., & Karlstrom, A. (2013). A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological, 56*, 70-80. doi:http://dx.doi.org/10.1016/j.trb.2013.07.012

Frejinger, E. (2008). Route choice analysis: data, models, algorithms and applications. *Unpublished PhD thesis, École Polytechnique Federale de Lausanne*.

Frejinger, E., & Bierlaire, M. (2010). On path generation algorithms for route choice models. *SH Daly and A.(eds.) Choice Modelling: The State-of-the-Art and the State-of-Practice*, 307-315.

Frejinger, E., Bierlaire, M., & Ben-Akiva, M. (2009). Sampling of alternatives for route choice modeling. *Transportation Research Part B: Methodological, 43*(10), 984-994.

Fu, X., & Juan, Z. (2016). Empirical analysis and comparisons about time-allocation patterns across segments based on mode-specific preferences. *Transportation, 43*(1), 37-51. doi:10.1007/s11116-014-9561-2

Gan, H., & Chen, S. (2013). Why Do Drivers Change Routes? Impact of Graphical Route Information Panels. *ITE Journal*.

Gan, H. C., Bai, Y., & Wei, J. (2013). Why do people change routes? Impact of information services. *INDUSTRIAL MANAGEMENT AND DATA SYSTEMS, 113*(3), 403-422.

Gao, S., Frejinger, E., & Ben-Akiva, M. (2010). Adaptive route choices in risky traffic networks: A prospect theory approach. *Transportation research part C: emerging technologies, 18*(5), 727-740.

Gärling, T., Gillholm, R., & Gärling, A. (1998). Reintroducing attitude theory in travel behavior research: The validity of an interactive interview procedure to predict car use. *Transportation, 25*(2), 129-146. doi:10.1023/a:1005004311776

Geda, A. (2014). *Effect of Choice Set Composition on Route Choice Models.* (M.Sc.), University of Florida.

Geller, R. (2009). *Four types of cyclists*: PortlandOnline.

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and psychological measurement, 55*(3), 377-393.

Golledge, R. G., Smith, T. R., Pellegrino, J. W., Doherty, S., & Marshall, S. P. (1985). A conceptual model and empirical analysis of children's acquisition of spatial knowledge. *Journal of Environmental Psychology, 5*(2), 125-152.

Greaves, S., Ellison, A., Ellison, R., Rance, D., Standen, C., Rissel, C., & Crane, M. (2015). A web-based diary and companion smartphone app for travel/activity surveys. *Transportation Research Procedia, 11*, 297-310.

Guevara, C. A., & Ben-Akiva, M. E. (2013). Sampling of alternatives in Multivariate Extreme Value (MEV) models. *Transportation Research Part B: Methodological, 48*, 31-52.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika, 19*(2), 149-161. doi:10.1007/bf02289162

Habib, K. N., Morency, C., Trépanier, M., & Salem, S. (2013). Application of an independent availability logit model (IAL) for route choice modelling: Considering bridge choice as a key determinant of selected routes for commuting in Montreal. *Journal of Choice Modelling, 9*, 14-26. doi:http://dx.doi.org/10.1016/j.jocm.2013.12.002

Hair, J. F. (2009). Multivariate data analysis.

Henn, V. (2003). Route Choice Making Under Uncertainty: a Fuzzy Logic Based Approach. In J.-L. Verdegay (Ed.), *Fuzzy Sets Based Heuristics for Optimization* (Vol. 126, pp. 277-292): Springer Berlin Heidelberg.

Hensher, D. A. (1994). Stated preference analysis of travel choices: the state of practice. *Transportation, 21*(2), 107-133.

Hess, S., & Daly, A. J. (2010). *Choice Modelling: The State-of-the-Art and the State-of-Practice: Proceedings from the Inaugural International Choice Modelling Conference: The State-of-the-art and the State-of-practice: Proceedings from the Inaugural International Choice Modelling Conference*: Emerald Group Publishing.

Hess, S., Quddus, M., Rieser-Schüssler, N., & Daly, A. (2015). Developing advanced route choice models for heavy goods vehicles using GPS data. *Transportation Research Part E: Logistics and Transportation Review, 77*, 29-44.

Higgs, B., & Abbas, M. (2013). *A two-step segmentation algorithm for behavioral clustering of naturalistic driving styles.* Paper presented at the 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013).

The High Cost of Congestion in Canadian Cities. (2012). Retrieved from http://www.google.ca/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0CDIQFjAB&url=http%3A%2F%2Fwww.mtq.gouv.qc.ca%2Fportal%2Fpage%2Fportal%2FLibrairie%2FPublications%2Ffr%2Fregions%2Fmontreal%2Fetude_eval_couts_congestion_mtl.pdf&ei=ikt0UtPyDImayQH5lYHoDA&usg=AFQjCNEH5DufEXVP69n6v4xDfBOHBjpCYw&sig2=eZ3JplhuiR86STQCLWpPaA&bvm=bv.55819444,d.aWc&cad=rja.

Hirtle, S. C., & Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory & cognition, 13*(3), 208-217.

Holding, C. S. (1994). Further evidence for the hierarchical representation of spatial information. *Journal of Environmental Psychology, 14*(2), 137-147. doi:http://dx.doi.org/10.1016/S0272-4944(05)80167-7

Hoogendoorn-Lanser, S. (2005). *Modeling Travel Behavior In Multi-Modal Networks.* (Doctorate), Technical University of Delft, Netherlands.

Hoogendoorn-Lanser, S., van Ness, R., & Bovy, P. (2005). *Path Size and Overlap in Multi-modal Transport Networks.* Paper presented at the 16th International Symposium on Transportation and Traffic Theory.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185. doi:10.1007/bf02289447

Horowitz, J. L., & Louviere, J. J. (1995). What is the role of consideration sets in choice modeling? *International Journal of Research in Marketing, 12*(1), 39-54.

Hurtubia, R., Nguyen, M. H., Glerum, A., & Bierlaire, M. (2014). Integrating psychometric indicators in latent class choice models. *Transportation Research Part A: Policy and Practice, 64*, 135-146. doi:https://doi.org/10.1016/j.tra.2014.03.010

Iida, Y., Uno, N., & Yamada, T. (1994, 31 Aug-2 Sep 1994). *Experimental analysis approach to analyze dynamic route choice behavior of driver with travel time information.* Paper presented at the Vehicle Navigation and Information Systems Conference, 1994. Proceedings., 1994.

Jan, O., Horowitz, A., & Peng, Z.-R. (2000). Using GPS Data to Understand Variations in Path Choice. *Transportation Research Record: Journal of the Transportation Research Board, 1725*(-1), 37-44. doi:10.3141/1725-06

Jensen, M. (1999). Passion and heart in transport—a sociological analysis on transport behaviour. *Transport Policy, 6*(1), 19-33.

Jou, R.-C., & Yeh, Y.-C. (2013). Freeway passenger car drivers' travel choice behaviour in a distance-based toll system. *Transport Policy, 27*(0), 11-19. doi:http://dx.doi.org/10.1016/j.tranpol.2012.12.005

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263-291.

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and psychological measurement*.

Kamargianni, M., Dubey, S., Polydoropoulou, A., & Bhat, C. (2015). Investigating the subjective and objective factors influencing teenagers' school travel mode choice – An integrated choice and latent variable model. *Transportation Research Part A: Policy and Practice, 78*, 473-488. doi:https://doi.org/10.1016/j.tra.2015.06.011

Kang, C., & Qin, K. (2016). Understanding operation behaviors of taxicabs in cities by matrix factorization. *Computers, Environment and Urban Systems, 60*, 79-88. doi:http://dx.doi.org/10.1016/j.compenvurbsys.2016.08.002

Kaplan, S., & Prato, C. G. (2012). Closing the gap between behavior and models in route choice: The role of spatiotemporal constraints and latent traits in choice set formation. *Transportation Research Part F: traffic psychology and behaviour, 15*(1), 9-24.

Kazagli, E., & Bierlaire, M. (2015). A Route Choice Model based on Mental Repre-sentations.

Kim, J., Sung, S., Namgung, M., & Jang, Y. (2005). *Development of dynamic route choice behavioral applied intelligent system theory*. Paper presented at the Proceedings of the Eastern Asia Society for Transportation Studies.

Koller-Matschke, I., Belzner, H., & Glas, R. (2013). Traffic Information and Individual Driver Behaviour *Proceedings of the FISITA 2012 World Automotive Congress* (Vol. 200, pp. 179-185): Springer Berlin Heidelberg.

Kroesen, M., & Handy, S. (2014). The relation between bicycle commuting and non-work cycling: results from a mobility panel. *Transportation, 41*(3), 507-527. doi:10.1007/s11116-013-9491-4

Kusakabe, T., Sharyo, T., & Asakura, Y. (2012). Effects of Traffic Incident Information on Drivers' Route Choice Behaviour in Urban Expressway Network. *Procedia - Social and Behavioral Sciences, 54*(0), 179-188. doi:http://dx.doi.org/10.1016/j.sbspro.2012.09.737

Lacombe, A., & Morency, C. (2016). *Modeling Taxi Trip Generation Using GPS Data: The Montreal Case*. Paper presented at the Transportation Research Board 95th Annual Meeting.

Lai, X., & Bierlaire, M. (2015). Specification of the cross-nested logit model with sampling of alternatives for route choice models. *Transportation Research Part B: Methodological, 80*, 220-234. doi:http://dx.doi.org/10.1016/j.trb.2015.07.005

Lebart, L., PIRON, A., Lebart, M., Morineau, A., & Piron, M. (2000). *Statistique exploratoire multidimensionnelle*: Dunod.

Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical assessment, research & evaluation, 12*(2), 1-11.

Levinson, D. (2003). The value of advanced traveler information systems for route choice. *Transportation research part C: emerging technologies, 11*(1), 75-87. doi:http://dx.doi.org/10.1016/S0968-090X(02)00023-2

Li, D., Miwa, T., Morikawa, T., & Liu, P. (2016). Incorporating observed and unobserved heterogeneity in route choice analysis with sampled choice sets. *Transportation research part C: emerging technologies, 67*(Supplement C), 31-46. doi:https://doi.org/10.1016/j.trc.2016.02.002

Li, H. (2004). Using Global Positioning System data to understand day-to-day dynamics of morning commute behavior. *Transportation research record.*(1895), 78-84.

Li, H., Guensler, R., & Ogle, J. (2005). Analysis of morning commute route choice patterns using Global Positioning System-based vehicle activity data. *Transportation research record.*(1926), 162-170.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.

Liu, L., Andris, C., & Ratti, C. (2010). Uncovering cabdrivers' behavior patterns from their digital traces. *Computers, Environment and Urban Systems, 34*(6), 541-548.

Lucidi, F., Giannini, A. M., Sgalla, R., Mallia, L., Devoto, A., & Reichmann, S. (2010). Young novice driver subtypes: relationship to driving violations, errors and lapses. *Accident Analysis & Prevention, 42*(6), 1689-1696.

Luisa De Maio, M., & Vitetta, A. (2015). Route choice on road transport system: A fuzzy approach. *Journal of Intelligent & Fuzzy Systems, 28*(5).

Lynch, K. (1960). *The image of the city* (Vol. 11): MIT press.

Mai, T. (2016). A method of integrating correlation structures for a generalized recursive route choice model. *Transportation Research Part B: Methodological, 93*, 146-161.

Mai, T., Bastin, F., & Frejinger, E. (2016). A decomposition method for estimating recursive logit based route choice models. *EURO Journal on Transportation and Logistics*, 1-23.

Mai, T., Bastin, F., & Frejinger, E. (2017). On the similarities between random regret minimization and mother logit: The case of recursive route choice models. *Journal of Choice Modelling, 23*, 21-33.

Mai, T., Fosgerau, M., & Frejinger, E. (2015). A nested recursive logit model for route choice analysis. *Transportation Research Part B: Methodological, 75*, 100-112.

Maio, M. L. D., Vitetta, A., & Watling, D. (2013). Influence of Experience on Users' Behaviour: A Day-to-Day Model for Route Choice Updating. *Procedia - Social and Behavioral Sciences, 87*(0), 60-74. doi:http://dx.doi.org/10.1016/j.sbspro.2013.10.594

Manley, E., Addison, J., & Cheng, T. (2015a). Shortest path or anchor-based route choice: a large-scale empirical analysis of minicab routing in London. *Journal of Transport Geography, 43*, 123-139.

Manley, E., Orr, S., & Cheng, T. (2015). A heuristic model of bounded route choice in urban areas. *Transportation research part C: emerging technologies, 56*, 195-209.

Manley, E. J., Addison, J. D., & Cheng, T. (2015b). Shortest path or anchor-based route choice: a large-scale empirical analysis of minicab routing in London. *Journal of Transport Geography, 43*, 123-139. doi:http://dx.doi.org/10.1016/j.jtrangeo.2015.01.006

Mannering, F., Kim, S.-G., Barfield, W., & Ng, L. (1994). Statistical analysis of commuters' route, mode, and departure time flexibility. *Transportation research part C: emerging technologies, 2*(1), 35-47. doi:http://dx.doi.org/10.1016/0968-090X(94)90018-3

Mannering, F. L. (1989). Poisson analysis of commuter flexibility in changing routes and departure times. *Transportation Research Part B: Methodological, 23*(1), 53-60. doi:http://dx.doi.org/10.1016/0191-2615(89)90023-4

Manski, C. F. (1977). The structure of random utility models. *Theory and decision, 8*(3), 229-254.

Marengo, D., Settanni, M., & Vidotto, G. (2012). Drivers' subtypes in a sample of Italian adolescents: relationship between personality measures and driving behaviors. *Transportation Research Part F: traffic psychology and behaviour, 15*(5), 480-490.

Mark, T., & Swait, J. (2008). Using stated preference and revealed preference data fusion modelling in health care. *Using Discrete Choice Experiments to Value Health and Health Care*, 217-234.

McFadden, D. (1978). *Modelling the choice of residential location*: Institute of Transportation Studies, University of California.

McFadden, D. (1986). The Choice Theory Approach to Market Research. *Marketing Science, 5*(4), 275-297.

McFadden, D. (1999). Rationality for Economists? *Journal of Risk and uncertainty, 19*(1), 73-105. doi:10.1023/a:1007863007855

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of applied Econometrics*, 447-470.

McFadden, D. L. (1984). Econometric analysis of qualitative response models. *Handbook of econometrics, 2*, 1395-1457.

McGowen, P. T. (2006). *Predicting Activity Types from GPS and GIS Data.* (Ph.D.), University of California Irvine.

Metrics, C. (2012). Ngene 1.1. 1 User Manual & Reference Guide. *Sydney, Australia: ChoiceMetrics*.

Metrolinx. (2008). *Costs of Road Congestion in the Greater Toronto and Hamilton Area: Impact and Cost Benefit Analysis of the Metrolinx Draft Regional Transportation Plan*. http://www.google.ca/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCsQFjAA&url=http%3A%2F%2Fwww.comt.ca%2Fenglish%2Futtf-congestion-2012.pdf&ei=L050UveBFNKEygHJhYHYCQ&usg=AFQjCNH68X65lizDzTrKX81TzvrAvYOZKQ&sig2=rNHvPAafb16ZtNIDZpZxog&bvm=bv.55819444,d.aWc.

Ministère des Transports du Québec. (2009). *Évaluation des coûts de la congestion routière dans la région de Montréal pour les conditions de référence de 2003*.

Mooi, E., & Sarstedt, M. (2010). *Cluster analysis*: Springer.

Morency, C., Trepanier, M., & Agard, B. (2011). *Typology of carsharing members.* Paper presented at the Transportation Research Board 90th Annual Meeting.

Morikawa, T. (1989). *Incorporating stated preference data in travel demand analysis.* (Ph.D.), Massachusetts Institute of Technology. Retrieved from http://hdl.handle.net/1721.1/14326

Muñoz, B., Monzon, A., & López, E. (2016). Transition to a cyclable city: Latent variables affecting bicycle commuting. *Transportation Research Part A: Policy and Practice, 84*, 4-17. doi:https://doi.org/10.1016/j.tra.2015.10.006

Murat, Y. S., & Uludag, N. (2008). Route choice modelling in urban transportation networks using fuzzy logic and logistic regression methods. *Journal of Scientific and Industrial research, 67*(1), 19.

Natural Resources Canada. (2009). *Canadian Vehicle Survey: 2009 Summary report*. Ottawa ON.

Nicolaas, G. (2011). Survey Paradata: a review.

Okamoto, K., Utsumi, A., Yamazoe, H., Miyashita, T., Abe, S., Takahashi, K., & Hagita, N. (2011). *Classification of Pedestrian Behavior in a Shopping Mall based on LRF and Camera Observations.* Paper presented at the MVA.

Ortúzar, J., & Willumsen, L. G. (2011). *Modelling transport*: Wiley Chichester:.

Ory, D. T., & Mokhtarian, P. L. (2005). Don't work, work at home, or commute? Discrete choice models of the decision for San Francisco Bay Area residents.

Pagès, J. (2004). Analyse factorielle de données mixtes. *Revue de statistique appliquée, 52*(4), 93-111.

Papinski, D. (2010). *Investigating route choice decisions using GPS and prompted-recall diary data.* (Doctoral dissertation), McMaster University.

Papinski, D. (2011). *Investigating Route Choice Decisions Using GPS and Prompted-recall Diary Data.* (PhD), McMaster University.

Papinski, D., & Scott, D. M. (2011). *Modeling Home-to-Work Route Choice Decisions Using GPS Data: A Comparison of Two Approaches for Generating Choice Sets.* Paper presented at the Transportation Research Board 90th Annual Meeting.

Papinski, D., Scott, D. M., & Doherty, S. T. (2009). Exploring the route choice decision-making process: A comparison of planned and observed routes obtained using person-based GPS. *Transportation Research Part F: traffic psychology and behaviour, 12*(4), 347-358. doi:http://dx.doi.org/10.1016/j.trf.2009.04.001

Parkany, E., Du, J., Aultman-Hall, L., & Gallagher, R. (2006a). Modeling stated and revealed route choice: consideration of consistency, diversion, and attitudinal variables. *Transportation Research Record: Journal of the Transportation Research Board*(1985), 29-39.

Parkany, E., Du, J., Aultman-Hall, L., & Gallagher, R. (2006b). Modeling Stated and Revealed Route Choice: Consideration of Consistency, Diversion, and Attitudinal Variables. *Transportation Research Record: Journal of the Transportation Research Board, 1985*(-1), 29-39. doi:citeulike-article-id:9038982 doi: 10.3141/1985-04

Parvaneh, Z., Arentze, T., & Timmermans, H. (2012). Understanding Travelers' Behavior in Provision of Travel Information: A Bayesian Belief Approach. *Procedia - Social and Behavioral Sciences, 54*(0), 251-260. doi:http://dx.doi.org/10.1016/j.sbspro.2012.09.744

Peeta, S., Ramos, J., & Pasupathy, R. (2000). Content of Variable Message Signs and On-Line Driver Behavior. *Transportation Research Record: Journal of the Transportation Research Board, 1725*(-1), 102-108. doi:10.3141/1725-14

Pele, N., & Morency, C. (2014). *When, Where and How Taxis Are Used in Montreal.* Paper presented at the Transportation Research Board 93rd Annual Meeting.

Peters, T., Adamowicz, W. L., & Boxall, P. C. (1995). Influence of choice set considerations in modeling the benefits from improved water quality. *Water Resources Research, 31*(7), 1781-1787.

Phaneuf, D. J., Taylor, L. O., & Braden, J. B. (2013). Combining revealed and stated preference data to estimate preferences for residential amenities: A GMM approach. *Land Economics, 89*(1), 30-52.

Poucin, G., Farooq, B., & Patterson, Z. (2016). *Pedestrian Activity Pattern Mining in WiFi-Network Connection Data.* Paper presented at the Transportation Research Board 95th Annual Meeting.

Prato, C. (2009a). Route choice modeling: past, present and future research directions. *Journal of Choice Modelling, 2*(1), 65-100. doi:http://dx.doi.org/10.1016/S1755-5345(13)70005-8

Prato, C., & Bekhor, S. (2006). Applying branch-and-bound technique to route choice set generation. *Transportation Research Record: Journal of the Transportation Research Board*(1985), 19-28.

Prato, C., & Bekhor, S. (2007a). Modeling route choice behavior: how relevant is the composition of choice set? *Transportation Research Record: Journal of the Transportation Research Board*.

Prato, C., & Bekhor, S. (2007b). Modeling route choice behavior: how relevant is the composition of choice set? *Transportation Research Record: Journal of the Transportation Research Board*(2003), 64-73.

Prato, C. G. (2009b). Route choice modeling: past, present and future research directions. *Journal of Choice Modelling, 2*(1), 65-100. doi:https://doi.org/10.1016/S1755-5345(13)70005-8

Prato, C. G. (2014). Expanding the applicability of random regret minimization for route choice analysis. *Transportation, 41*(2), 351-375.

Prato, C. G., Bekhor, S., & Pronello, C. (2012). Latent variables and route choice behavior. *Transportation, 39*(2), 299-319.

Quattrone, A., & Vitetta, A. (2011). Random and fuzzy utility models for road route choice. *Transportation Research Part E: Logistics and Transportation Review, 47*(6), 1126-1139. doi:http://dx.doi.org/10.1016/j.tre.2011.04.007

Rakotomalala, R. (2005). *TANAGRA: un logiciel gratuit pour l'enseignement et la recherche.* Paper presented at the EGC.

Ramaekers, K., Reumers, S., Wets, G., & Cools, M. (2013). Modelling Route Choice Decisions of Car Travellers Using Combined GPS and Diary Data. *Networks and Spatial Economics, 13*(3), 351-372. doi:10.1007/s11067-013-9184-8

Ramming, M. S. (2002). *Network knowledge and route choice.* (Doctoral dissertation), Massachusetts Institute of Technology.

Raveau, S., Álvarez-Daziano, R., Yáñez, M., Bolduc, D., & de Dios Ortúzar, J. (2010). Sequential and simultaneous estimation of hybrid discrete choice models: some new findings. *Transportation Research Record: Journal of the Transportation Research Board*(2156), 131-139.

Reinoso, G., & Farooq, B. (2015). *URBAN PULSE ANALYSIS USING BIG DATA*. Paper presented at the 50th Annual Conference of Canadian Transportation Research Forum.

Richardson, A. J., Ampt, E. S., & Meyburg, A. H. (1995). *Survey methods for transport planning*: Eucalyptus Press Melbourne.

Ridwan, M. (2004). Fuzzy preference based traffic assignment problem. *Transportation research part C: emerging technologies, 12*(3), 209-233. doi:https://doi.org/10.1016/j.trc.2004.07.003

Rieser-Schüssler, N., Balmer, M., & Axhausen, K. W. (2013). Route choice sets for very high-resolution data. *Transportmetrica A: Transport Science, 9*(9), 825-845.

Rose, J. M., & Bliemer, M. C. J. (2008). Stated Preference Experimental Design Strategies. In D. A. Hensher & K. J. Button (Eds.), *Handbook of Transportation Modeling* (Second ed.). Amsterdam, The Netherlands: Elsevier.

Sanko, N. (2002). Guidelines for State Preference Experiment Design: Ecole Nationale des Pont et Chaussees, Paris.

Saporta, G. (2003). A control chart approach to select eigenvalues in principal component and correspondence analysis. *54th Session of the International Statistical Institute-Berlin*.

Sarkar, P. P., & Mallikarjuna, C. (2017). Effect of perception and attitudinal variables on mode choice behavior: A case study of Indian city, Agartala. *Travel Behaviour and Society*. doi:https://doi.org/10.1016/j.tbs.2017.04.003

Saunier, N., Mourji, N., & Agard, B. (2011). Mining microscopic data of vehicle conflicts and collisions to investigate collision factors. *Transportation Research Record: Journal of the Transportation Research Board*(2237), 41-50.

Schlaich, J. (2010). Analyzing route choice behavior with mobile phone trajectories. *Transportation Research Record: Journal of the Transportation Research Board*(2157), 78-85.

Schuessler, N., & Axhausen, K. (2009). *Accounting for Route Overlap in Urban and Sub-urban Route Choice Decisions Derived from GPS Observations*. Paper presented at the 12th International conference on travel behavior research, Jaipur, India.

Shaaban, K., & Kim, I. (2016). Assessment of the taxi service in Doha. *Transportation Research Part A: Policy and Practice, 88*, 223-235. doi:http://dx.doi.org/10.1016/j.tra.2016.04.011

Sikka, N. (2012). *Understanding travelers' route choice behavior under uncertainty.* (Ph.D.), University of Iowa.

Srinivasan, S., & Dhakar, N. S. (2013). *Route-choice Modeling using GPS-based Travel Surveys*. Retrieved from

Sun, H., Wu, J., Ma, D., & Long, J. (2014). Spatial distribution complexities of traffic congestion and bottlenecks in different network topologies. *Applied Mathematical Modelling, 38*(2), 496-505. doi:http://dx.doi.org/10.1016/j.apm.2013.06.027

Swait, J., & Ben-Akiva, M. (1987a). Empirical test of a constrained choice discrete model: mode choice in Sao Paulo, Brazil. *Transportation Research Part B: Methodological, 21*(2), 103-115.

Swait, J., & Ben-Akiva, M. (1987b). Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological, 21*(2), 91-102.

Tawfik, A. M., & Rakha, H. (2013). *A Latent Class Choice Model of Heterogeneous Drivers Route Choice Behavior Based on a Real-World Experiment.* Paper presented at the 20th International Symposium on Transportation and Traffic Theory, Noordwijk, Netherlands.

Tawfik, A. M., Rakha, H. A., & Miller, S. D. (2010a, 21-24 June 2010). *Driver route choice behavior: Experiences, perceptions, and choices.* Paper presented at the Intelligent Vehicles Symposium (IV), 2010 IEEE.

Tawfik, A. M., Rakha, H. A., & Miller, S. D. (2010b, 19-22 Sept. 2010). *An experimental exploration of route choice: Identifying drivers choices and choice patterns, and capturing network evolution.* Paper presented at the 13th International IEEE Conference.

Tawfik, A. M., Rakha, H. A., & Miller, S. D. (2010c). *An experimental exploration of route choice: identifying drivers choices and choice patterns, and capturing network evolution.* Paper presented at the Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on.

Tawfik, A. M., Szarka, J., House, L., & Rakha, H. A. (2011a, 5-7 Oct. 2011). *Disaggregate route choice models based on driver learning patterns and network experience.* Paper presented at the 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC).

Tawfik, A. M., Szarka, J., House, L., & Rakha, H. A. (2011b, 5-7 Oct. 2011). *Disaggregate route choice models based on driver learning patterns and network experience.* Paper presented at the Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on.

Tian, L., Huang, H., & Liu, T. (2010). Day-To-Day Route Choice Decision Simulation Based on Dynamic Feedback Information. *Journal of Transportation Systems Engineering and Information Technology, 10*(4), 79-85. doi:http://dx.doi.org/10.1016/S1570-6672(09)60056-8

Train, K. E. (2009). *Discrete choice methods with simulation*: Cambridge university press.

Transport, A. M. d. (2013). *Enquête Origine-Destination 2013: La mobilité des personnes dans la région de Montréal-FAITS SAILLANTS*. Montreal Retrieved from https://rtm.quebec/Media/Default/pdf/section8/enquete-od-2013-faits-saillants.pdf.

Turcotte, M. (2008). Life in metropolitan areas *Dependence on cars in urban neighbourhoods*. Retrieved from http://www.statcan.gc.ca/pub/11-008-x/2008001/article/10503-eng.htm#1

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty, 5*(4), 297-323.

Ulleberg, P. (2001). Personality subtypes of young drivers. Relationship to risk-taking preferences, accident involvement, and response to a traffic safety campaign. *Transportation Research Part F: traffic psychology and behaviour, 4*(4), 279-297. doi:http://dx.doi.org/10.1016/S1369-8478(01)00029-8

Vacca, A., & Meloni, I. (2014). *Understanding route choice behavior: an analysis using GPS based data*. Paper presented at the Transport Research Arena (TRA) 5th Conference: Transport Solutions from Research to Deployment.

Vitetta, A. (2016). A quantum utility model for route choice in transport systems. *Travel Behaviour and Society, 3*, 29-37. doi:http://dx.doi.org/10.1016/j.tbs.2015.07.003

Vovsha, P., & Bekhor, S. (1998). The link-nested logit model of route choice: overcoming the route overlapping problem. *Transportation Research Record: Journal of the Transportation Research Board, 1645*, 133-142.

Vredin Johansson, M., Heldt, T., & Johansson, P. (2006). The effects of attitudes and personality traits on mode choice. *Transportation Research Part A: Policy and Practice, 40*(6), 507-525. doi:https://doi.org/10.1016/j.tra.2005.09.001

Walker, J., & Ben-Akiva, M. (2002). Generalized random utility model. *Mathematical Social Sciences, 43*(3), 303-343. doi:https://doi.org/10.1016/S0165-4896(02)00023-9

Walker, J., Ben-Akiva, M., & Bolduc, D. (2004). *Identification of the logit kernel (or mixed logit) model*. Paper presented at the 10th International Conference on Travel Behavior Research, Lucerne, Switzerland, August.

Walker, J. L. (2001). *Extended discrete choice models: integrated framework, flexible error structures, and latent variables*. Massachusetts Institute of Technology.

Wardman, M., Bonsall, P. W., & Shires, J. D. (1997). Driver response to variable message signs: a stated preference investigation. *Transportation research part C: emerging technologies, 5*(6), 389-405. doi:http://dx.doi.org/10.1016/S0968-090X(98)00004-7

Wen, C.-H., & Lai, S.-C. (2010). Latent class models of international air carrier choice. *Transportation Research Part E: Logistics and Transportation Review, 46*(2), 211-221.

White, C. E., Bernstein, D., & Kornhauser, A. L. (2000). Some map matching algorithms for personal navigation assistants. *Transportation research part C: emerging technologies, 8*(1), 91-108.

Wiener, J. M., & Mallot, H. A. (2003). 'Fine-to-coarse'route planning and navigation in regionalized environments. *Spatial cognition and computation, 3*(4), 331-358.

Wolf, J. (2004). *Applications of new technologies in travel surveys*. Paper presented at the 7th International Conference on Travel Survey Methods, Costa Rica.

Woo, C. K., Cheng, Y. S., Li, R., Shiu, A., Ho, S. T., & Horowitz, I. (2015). Can Hong Kong price-manage its cross-harbor-tunnel congestion? *Transportation Research Part A: Policy and Practice, 82*, 94-109. doi:http://dx.doi.org/10.1016/j.tra.2015.09.002

Xu, H., Zhou, J., & Xu, W. (2011). A decision-making rule for modeling travelers' route choice behavior based on cumulative prospect theory. *Transportation research part C: emerging technologies, 19*(2), 218-228. doi:http://dx.doi.org/10.1016/j.trc.2010.05.009

Yang, H., Ye, M., Tang, W. H., & Wong, S. C. (2005). Regulating taxi services in the presence of congestion externality. *Transportation Research Part A: Policy and Practice, 39*(1), 17-40.

Yazdizadeh, A. (2016). *A Generic Form for Capturing Unobserved Heterogeneity in Discrete Choice Modelling: Application to Neighborhood Location Choice.* Concordia University.

Yu, J., & Peeta, S. (2011). Experimental analysis of a hybrid route choice model to capture dynamic behavioral phenomena under advanced information systems. *KSCE Journal of Civil Engineering, 15*(1), 175-185. doi:10.1007/s12205-011-0947-9

Zhou, J., & Golledge, R. (2006). A three-step general map matching method in the GIS environment: Travel/transportation study perspective. *International Journal of Geographical Information System, 8*(3), 243-260.

Zimmermann, M., Mai, T., & Frejinger, E. (2017a). Bike route choice modeling using GPS data without choice sets of paths. *Transportation research part C: emerging technologies, 75*, 183-196.

Zimmermann, M., Mai, T., & Frejinger, E. (2017b). Bike route choice modeling using GPS data without choice sets of paths. *Transportation Research Part C: Emerging Technologies, 75*(Supplement C), 183-196. doi:https://doi.org/10.1016/j.trc.2016.12.009